arXiv:2602.02364v1 [quant-ph] 2 Feb 2026

# 通过理论机制保障混合量子学习中的隐私

**Hoang M. Ngo** [1] **Tre' R. Jeter** [1 2] **Incheol Shin** [3] **Wanli Xing** [4] **Tamer Kahveci** [1] **My T. Thai** [1]

**摘要**

量子机器学习（QML）凭借提升分类等经典机器学习（ML）任务的潜力，正日益普及。尽管量子噪声通常被视为量子计算的一大挑战，但它也为增强隐私提供了独特机会。具体而言，固有的量子噪声是一种天然的随机资源；在差分隐私（DP）框架内对其进行严格分析，并将其与经典机制组合后，即可满足形式化的（$\varepsilon$，$\delta$）-差分隐私保证。这使得在不影响隐私预算的前提下减少所需的经典扰动成为可能，从而有望提升模型效用。然而，利用经典噪声和量子噪声相结合来实现隐私保护，仍未得到探索。在本研究中，我们提出混合加噪机制 HYPER-Q，将经典噪声与量子噪声相结合，以保护量子机器学习模型的隐私。我们全面分析了其隐私保证，并确立了其效用的理论界。通过实证研究，我们证明，与现有基于经典噪声的机制相比，HYPER-Q 在多个真实世界数据集上的对抗鲁棒性更强。

## 1. 引言

量子机器学习（QML）是一种新兴且极具吸引力的范式，它融合了量子系统的计算优势与机器学习（ML）的建模能力。量子系统的一个基本特征是量子噪声，即固有的随机性，以及与环境相互作用所导致的退相干。尽管量子噪声通常被视为实现容错量子计算的一大障碍，但它也提供了机会，可作为隐私保护的天然内在随机性来源。

在经典机器学习中，差分隐私（DP）（Dwork，2006年）已成为提供形式化隐私保证的标准框架。差分隐私确保当某个个体的数据被添加到输入数据集或从中移除时，算法的输出不会发生显著变化，从而保护个体隐私。除了在隐私保护中的作用外，差分隐私还被扩展用于认证机器学习模型对对抗攻击的鲁棒性（Lecuyer 等人，2019年；Cohen等人，2019年）。差分隐私通常通过在学习过程中注入经过精细校准的随机噪声（如高斯噪声或拉普拉斯噪声）来实现（Geng &Viswanath，2012年；Balle & Wang，2018；Ji &Li，2024年）。此外，还可以通过子采样（Balle 等人，2018）、迭代组合（Feldman 等人，2018）以及基于扩散的机制（Balle 等人，2019年）等额外随机技术来增强整体隐私保证。然而，在任意组合随机技术的情况下，理论上的隐私放大并不能得到保证。

近期研究将差分隐私的概念扩展到量子领域，从而引出量子差分隐私（QDP）（Du 等人，2021a；Hirche 等人，2023年）。然而，若干关键挑战仍未得到解决。首先，现有工作主要侧重于为量子数据定义隐私保证。然而，大多数实用的近期量子机器学习应用都是混合模型，它们处理经典数据，并仅将量子电路用作中间处理组件。这种混合架构带来了一个关键的隐私挑战：仅对中间量子层应用差分隐私保证，并不能确保整个模型的端到端隐私，尤其是当前置经典组件较为敏感时。其次，经典噪声（例如高斯噪声、拉普拉斯噪声）与固有量子噪声之间的相互作用尚未得到研究。这一研究空白至关重要，因为某些类型的量子噪声（如去极化噪声）能够自然地向学习过程注入随机性，而不会显著降低模型性能（Du 等人，2021a）。这提出了一个关键

[1]佛罗里达大学计算机与信息科学及工程系，美国佛罗里达州盖恩斯维尔 [2]MITRE，美国弗吉尼亚州费尔法克斯 [3]釜庆大学人工智能融合系，韩国釜山 [4]佛罗里达大学教育系，美国佛罗里达州盖恩斯维尔。通讯作者：My T. Thai< mythai@cise.uf.edu>。

预印本。*2026*年*2*月*3*日。

开放性问题：这种固有的量子随机性能否被正式用作一种随机技术，以放大源自前置经典机制的隐私保证？迄今为止，尚无工作从理论上确立如何在这些混合模型中组合经典与量子噪声源的隐私保证。此外，理解这一关系对于控制预设隐私预算至关重要，尤其是考虑到物理设备中的量子噪声本质上是动态的，且难以精确控制。

**贡献**。本工作的关键贡献与见解可重点概括如下：

**（i）混合隐私保护机制**。我们提出 HYPER-Q，一种面向量子神经网络（QNNs）的混合隐私保护机制（**HY**brid **P**rivacy-pres**ER**ving mechanism for**Q**uantum）。据我们所知，这是首个研究经典噪声与量子噪声的联合效应、从而在量子混合模型中放大差分隐私的工作。具体而言，HYPER-Q 通过经典输入扰动（例如高斯噪声）建立基线隐私保证，随后由作为后处理操作的量子噪声信道在理论上加以放大，其中以退极化噪声为主要关注点。

**（ii）隐私保证分析**。我们对 HYPER-Q 的差分隐私保证进行了严格分析。我们的机制是一个复合 $Q^{(\eta)} \circ A$，其中 $A$ 是满足原始 $(\varepsilon, \delta)$-差分隐私的经典机制，$Q^{(\eta)}$ 是纳入退极化噪声信道的量子后处理操作，并作为我们的主要分析情形。我们分析该复合如何实现新的放大后隐私参数 $(\varepsilon', \delta')$。我们给出三项主要分析结果，其中两项针对退极化噪声，一项扩展至其他量子噪声信道。

- 第一（定理 4.1）：我们证明，在一个 a $d$ 维希尔伯特空间中，量子后处理可作为隐私放大器，通过严格降低失败概率来实现这一点（达到 $\delta' = \left[\eta\,(1-e^{\varepsilon})_{d} + (1-\eta)\delta\right]_{+} < \delta$），同时保持隐私损失不变（$\varepsilon' = \varepsilon$）。这直接意味着可认证对抗鲁棒性更强。

- 第二（定理 4.4）：我们证明，在特定条件下，可以同时放大两个参数 $\varepsilon'$ 和 $\delta'$。这项分析得出了两个关键见解。首先，我们说明了如何选择正算子值测度（POVM）以最大化隐私增益：当所有 POVM 元素的迹相等时，$\delta'$ 的上界最小（即保证最强）。其次，我们推导出了量子噪声 $\eta$ 必须超过的明确阈值，以确保严格放大 $\varepsilon'$ 和 $\delta'$。

- 第三（定理 4.7 和 4.9）：我们将针对退极化噪声的隐私分析扩展至其他非对称量子噪声信道，并指出量子曲棍球棒散度的收缩是隐私放大的根本机制。我们基于热弛豫（$\delta' = (2\sqrt{\eta} - \eta)\delta$），推导出广义振幅阻尼（GAD）下的严格隐私放大；此外，在采用乘积赤道编码的假设下，也推导出广义退相位（GD）下的严格隐私放大，其中相位相干性的抑制会将失败概率缩减至 $\delta' = |1 - 2\eta|\delta$。

**（iii）效用分析**。我们推导出一个形式化的效用界（定理 4.10），用于量化模型在去极化信道下的性能。具体而言，我们将总误差刻画为经典噪声方差（$\sigma$）与去极化因子（$\eta$）之间的高概率权衡。

**（iv）实证实验**。我们通过实证表明，在固定的端到端隐私预算下，HYPER-Q 在多个数据集上始终比标准的纯经典差分隐私机制具有更强的对抗鲁棒性。这些结果表明，用量子退极化噪声替代经典噪声可以在不削弱隐私保证的情况下获得更高的性能。

## 2. 预备知识

### 2.1. 量子信息基础

**量子比特与态**。量子计算系统作用于量子比特（量子比特）。与经典比特不同，量子比特可以处于 0 和 1 的叠加态。一个 $n$ 量子比特系统位于 $2^n$ 维希尔伯特空间 $\mathcal{H}$ 中。理想（纯）态由向量 $|\psi\rangle$ 表示，而一般（可能含噪声的）态则由密度矩阵 $\rho$ 描述：$d \times d$ 迹为一的半正定矩阵（即 $\mathrm{Tr}[\rho] = 1$）。

**量子信道**。量子态的演化（包括噪声效应）可用量子信道建模。例如，去极化信道记作 $f^{(\eta)}{}_{\text{dep}}$，它以概率 $\eta$ 将状态 $\rho$ 替换为最大混合态 $\frac{I}{d}$，并以概率 $1-\eta$ 使其保持不变：

$$f^{(\eta)}_{\text{dep}}(\rho) = (1-\eta)\rho + \eta\frac{I}{d}$$

其中，$\eta \in [0,1]$ 表示概率，$I$ 表示单位矩阵，$d$ 表示希尔伯特空间的维数。

通过测量从量子态中提取经典信息。一般测量由一组算子 $E_k$ 构成，称为正算子值测度 (POVM)。对于状态 $\rho$，观测到结果 $k$ 的概率为：

$$\Pr(\text{outcome} = k) = \mathrm{Tr}[E_k\rho].$$

### 2.2. 差分隐私

差分隐私 (DP) 提供了一种形式化保证：数据集中任何单个样本的存在与否，对输出的影响有限（Dwork，2006）。更正式地说：

**定义 2.1** ($(\varepsilon, \delta)$-差分隐私)。若随机机制 $\mathcal{M}: \mathcal{D} \to \mathcal{R}$ 满足 $(\varepsilon, \delta)$-差分隐私，则对于任意两个仅相差一个元素的相邻数据集 $D_1$ 和 $D_2$，以及任意输出子集 $S \subseteq \mathcal{R}$，以下不等式成立：

$$\Pr[\mathcal{M}(D_1) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(D_2) \in S] + \delta$$

，

其中，$\varepsilon \geq 0$ 是隐私损失参数，而 $\delta \in [0, 1)$ 是失败概率。$\varepsilon$ 越小或 $\delta$越小，隐私保护就越强。

差分隐私的一种等价刻画可通过**曲棍球棒散度**给出。对于两个分布 $P$和 $Q$，曲棍球棒散度定义如下：

$$\mathrm{D}_{e^{\varepsilon}}(P \| Q) = \int \max(0, P(x) - e^{\varepsilon} Q(x)) dx$$

当且仅当对于所有相邻的 $D_1, D_2$，$\mathrm{D}_{e^{\varepsilon}}(\mathcal{M}(D_1) \| \mathcal{M}(D_2)) \leq \delta$ 时，机制 $\mathcal{M}$ 满足 $(\varepsilon, \delta)$-差分隐私。
该框架可扩展至量子场景（Hirche 等人，2023年），其中，状态的量子 hockey-stick 散度 $\rho$、$\rho'$ 定义如下：

$$\mathrm{D}_{e^{\varepsilon}}^{(q)}(\rho \| \rho') = \mathrm{Tr}\left[(\rho - e^{\varepsilon} \rho')_{+}\right]$$

量子机制 $\mathcal{E}$ 满足（$\varepsilon$，$\delta$）-量子差分隐私，当且仅当对于任意相邻状态 $\rho$、$\rho'$，其散度有界于 $\delta$其中 $\mathrm{D}_{e^{\varepsilon}}^{(q)}(\mathcal{E}(\rho) \| \mathcal{E}(\rho')) \leq \delta$。
**加噪机制**。实现差分隐私的一种标准方法，是添加与函数敏感度成比例的噪声；函数敏感度指更改一个数据点所导致的最大输出变化。高斯机制向函数 $f: \mathcal{D} \to \mathbb{R}$ 添加噪声 $\eta_{\text{cdp}} \sim \mathcal{N}(0, \sigma^2 I)$，并依据该函数的 $L_2$ 敏感度：

$$\Delta_2(f) = \max_{D_1, D_2} ||f(D_1) - f(D_2)||_2$$

该机制的输出为 $f(x) + \eta_{\text{cdp}}$。选择适当的 $\sigma$后，该机制满足（$\varepsilon$，$\delta$）-差分隐私。关于混合量子机器学习的更多背景，以及差分隐私与对抗鲁棒性之间的联系，请参阅附录A。

## 3. 相关工作

**经典机器学习中的差分隐私**。差分隐私（DP）已被确立为一种领先的保护机器学习工作流中数据的框架。差分隐私提供了形式化保证（Dwork 等人，2006年），确保单个数据点的纳入或排除对算法的输出影响有限，从而最大限度地降低信息泄露风险。在机器学习中，实践中实现差分隐私最常见的方法是向学习过程注入经过校准的随机噪声。这种噪声可在不同阶段引入，例如扰动输入数据（Lecuyer 等人，2019年；Phan等人，2019年；Cohen 等人，2019年）、优化过程中的梯度（Abadi等人，2016年；Ghazi等人，2025年），或最终的模型参数（Yuan等人，2023年）。

输入扰动对于提供实例级隐私尤其有效，也是认证模型预测对抗鲁棒性的关键技术（Lecuyer 等人，2019；Cohen等人，2019）。标准机制（如高斯机制或拉普拉斯机制）会根据函数的敏感度缩放所添加的噪声，以提供 $(\varepsilon, \delta)$-差分隐私保证（Dwork & Roth，2014）。为缓解噪声注入通常导致的模型性能下降，一条重要的研究路线聚焦于隐私放大。其核心思想是，某些随机过程无需额外的初始噪声，即可增强最终的隐私保证。隐私放大还可以通过子采样（Bun 等人，2015年；Balle 等人，2018；Wang 等人，2019年；Koga 等人，2022年）、洗牌（Cheu 等人，2018；Er-lingsson 等人，2019年；Balle 等人，2019b）、迭代组合（Feldman 等人，2018）以及专门形式的后处理（Balle等人，2019a；Ye & Shokri，2022）等成熟技术实现。特别是，后处理具有基础性作用：尽管标准后处理绝不会削弱隐私保证（Dwork，2006），但某些随机变换能够主动增强隐私。然而，并非所有随机源的组合都能带来隐私放大。例如，对高斯机制进行额外的高斯变换后处理可以放大隐私，而将高斯机制与拉普拉斯变换组合则不会产生这种效果。

**量子情境下的差分隐私**。随着人们对隐私保护量子计算和量子机器学习（QML）的兴趣日益浓厚，差分隐私的概念近期已扩展至量子情境。其基础概念由（Zhou 和 Ying，2017）提出，他们给出的量子差分隐私定义是经典差分隐私的直接量子类比。在此基础上，（Du 等人，2021a）展示了其在量子机器学习中的实际应用，指出可利用量子固有噪声，在量子分类器中实现量子差分隐私。具体而言，他们分析了退极化噪声信道作为一种隐私保护机制的作用，并推导出噪声强度与相应的 $(\varepsilon, \delta)$-量子差分隐私保证之间的数学关系。他们还证明，这种隐私机制还能同时增

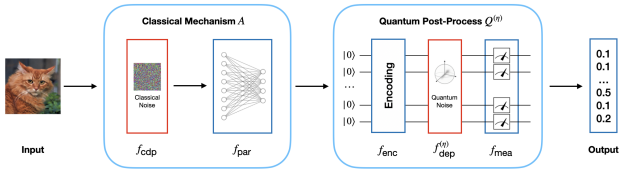



图 *1.* 所提出的 HYPER-Q 概述。

强模型的对抗鲁棒性。随后，（Hirche 等人，2023年）构建了一个全面的量子差分隐私理论框架。他们运用量子相对熵等工具，为量子差分隐私提供了更一般且严谨的理论基础。近期的研究（Watkins 等人，2023年；Bai 等人，2024年；Song 等人，2025年）考察了退极化、比特翻转和相位翻转等各种量子噪声源对量子差分隐私预算的影响。

尽管在纯量子系统或纯经典系统的隐私定义方面取得了这些进展，但对于近期开拓量子优势所必需的混合量子-经典架构，仍存在一个关键缺口。这些模型对于将量子计算应用于现实世界问题至关重要。然而，迄今为止，尚无研究从理论上确立如何在混合量子模型中组合经典与量子噪声源的隐私保证。这一重大缺口凸显了我们提出 HYPER-Q 的重要性，也表明有必要进一步探索混合方法，将传统差分隐私机制与量子系统固有的隐私属性相结合。

## 4. 混合加噪机制

本节将介绍一种保护隐私的机制，该机制结合经典噪声与量子噪声，以在量子神经网络模型中实现差分隐私（DP）。我们首先描述混合机制的结构，然后分析其差分隐私保证，最后给出效用界，以刻画噪声对模型性能的影响。

### 4.1. 机制概述

所提出的机制旨在从两个层面减轻隐私泄露。首先，经典数据在进入量子电路之前可能容易遭受重构攻击。为防止此类暴露，我们引入经典噪声机制来扰动输入。其次，我们利用量子退极化噪声的固有特性，在编码后增强隐私。研究表明，在测量次数趋于无限的理想情况下，这种噪声能够保留效用（Du 等人，2021b）。通过结合经典噪声与量子噪声，我们的双层方法降低了对过量经典噪声的依赖，从而在将效用损失降至最低的同时，实现更强的隐私保护。

我们使用基于函数的模块化表示形式，正式描述机制的各个阶段（见图 1）：

**经典噪声函数** $f_{\text{cdp}}$：$\mathbb{X} \to \mathbb{X}$。此函数向输入添加经过校准的经典噪声，以提供基准差分隐私保证。

$$f_{\text{cdp}}(x) = x + \eta_{\text{cdp}}, \quad \text{where } \eta_{\text{cdp}} \sim \mathcal{N}(0, \sigma^2 I)$$

这里，噪声 $\eta_{\text{cdp}}$ 服从协方差为 $\sigma^2 I$的多元高斯分布。

**参数化线性变换** $f_{\text{par}}$：$\mathbb{X} \to \mathbb{Y}$。此函数充当可学习的经典层，将输入数据转换到特征空间中。权重 $W$ 和偏置 $b$在模型训练过程中学习得到。

$$f_{\text{par}}(x') = Wx' + b = y$$

**量子编码函数** $f_{\text{enc}}$：$\mathbb{Y} \to \mathcal{H}$。此函数将经典特征向量 $y$ 编码为量子态 $\rho$，该量子态位于由 $n$ 个量子比特组成的 $d$维希尔伯特空间 $\mathcal{H}$ 中（$d = 2^n$）。设 $|\psi_y\rangle = \prod_{j=1}^{n} e^{-iy_j H_j}|0\rangle^{\otimes n}$为编码后的纯态向量，其中 $H_j$ 为厄米算子。输出为对应的密度矩阵：

$$f_{\text{enc}}(y) = |\psi_y\rangle\langle\psi_y| = \rho$$

**退极化噪声信道** $f^{(\eta)}_{\text{退极化}}$：$\mathcal{H} \to \mathcal{H}$。该 Quantum 信道通过直接对编码态 $\rho$施加噪声，添加了第二层随机性。后续分析将表明，这一过程会放大经典噪声层所提供的初始隐私保证。

$$f^{(\eta)}_{\text{dep}}(\rho) = (1-\eta)\rho + \eta\frac{I}{d} = \tilde{\rho}$$

其中，$\eta \in [0,1]$ 是退极化概率，$I$ 是作用于 $\mathcal{H}$上的单位算子。

**测量函数** $f_{\text{mea}}$：$\mathcal{H} \to \mathbb{Z}$。这一最终阶段将带噪量子态 $\tilde{\rho}$ 映射为输出空间 $\mathbb{Z} = \{0, 1, \ldots, K-1\}$中的单个经典类别标签 $z$。这种映射本质上具有随机性，其形式化定义如下：

$$\Pr(f_{\text{mea}}(\tilde{\rho}) = k) = \text{Tr}[E_k \tilde{\rho}], \quad \forall k$$

这种混合方法可以独立调节经典噪声和量子噪声，从而灵活实现隐私-效用权衡。其模块化设计还支持对隐私保证和性能影响进行理论分析，具体如下。

### 4.2. 差分隐私界

现在，我们定义差分隐私分析中使用的概念。具体而言，我们提出的机制可以表示为组合 $Q^{(\eta)} \circ A$，其中 $A = f_{\text{par}} \circ f_{\text{cdp}}$ 是满足 $(\varepsilon,\ \delta)$-差分隐私的经典机制，而 $Q^{(\eta)} = f_{\text{mea}} \circ f^{(\eta)}_{\text{退极化}} \circ f_{\text{enc}}$ 是由退极化噪声参数 $\eta$ 控制的量子后处理操作。假设随机过程

$f_{\text{cdp}}$ 满足 $(\varepsilon,\ \delta)$-差分隐私。根据后处理定理（Dwork，2006），机制 $A$ 也满足 $(\varepsilon,\ \delta)$-差分隐私。虽然我们主要关注退极化噪声，但该框架具有模块化特性。将 $f^{(\eta)}{}_{\text{退极化}}$ 替换为其他量子噪声信道，即可研究各种量子后处理效应。

我们的目标是分析复合机制如何获得新的隐私参数 $(\varepsilon',\ \delta')$，以及这些参数如何放大原有保证 $(\varepsilon,\delta)$。具体而言，我们给出两项针对退极化噪声的分析，随后第三项分析将这些结果推广到更广泛的量子噪声信道。在第一项分析中，我们表明 $Q^{(\eta)}\circ A$ 能够通过如下证明改善失败概率：$\varepsilon'=\varepsilon$ 以及 $\delta'<\delta$。在第二项分析中，我们证明在某些条件下，$Q^{(\eta)}\circ A$ 能够同时放大隐私损失与失败概率，从而实现 $\varepsilon'<\varepsilon$ 以及 $\delta'<\delta$。在第三项分析中，我们将这些结果从退极化噪声推广到更广泛的一类信道，包括广义振幅阻尼（GAD）和广义退相位（GD）。由此，我们确定任意量子噪声信道实现差分隐私放大所需的通用条件。所有证明与细节见附录B。

#### 4.2.1. 第一项分析——放大失败概率

我们研究在隐私损失参数 $\varepsilon$ 固定的情况下，量子后处理如何放大失败概率。定理 4.1 通过为复合机制 $Q^{(\eta)}\circ A$ 的失败概率 $\delta'$ 建立新的界来形式化这一结果，同时将隐私损失固定为$\varepsilon'=\varepsilon$。该定理的证明通过两个关键步骤衔接经典与量子散度度量：**(1)** 证明测量所得概率的经典曲棍球棒散度不超过测量前量子态的量子曲棍球棒散度；**(2)** 证明该量子散度在 $f^{(\eta)}{}_{\text{退极化}}$ 作用下会按因子 $(1-\eta)$ 收缩。

**定理 4.1**(失败概率放大)。 设$A:\ \mathbb{X}\to\mathcal{P}(\mathbb{Y})$ 是一个满足 $(\varepsilon,\delta)$-差分隐私的经典机制，其中 $A=f_{par}\circ f_{cdp}$；设 $Q^{(\eta)}:\ \mathbb{Y}\to\mathcal{P}(\mathbb{Z})$ 是在 $d$维希尔伯特空间中的量子机制，定义为 $Q^{(\eta)}=f_{mea}\circ f^{(\eta)}{}_{\text{退极化}}\circ f_{enc}$，其中 $0\le\eta\le 1$ 是退极化噪声因子。那么，复合机制 $Q^{(\eta)}\circ A$ 满足 $(\varepsilon',\delta')$-差分隐私，其中

$$\varepsilon'=\varepsilon,\quad \delta'=\left[\frac{\eta(1-e^{\varepsilon})}{d}+(1-\eta)\delta\right]_{+}$$

由此可知，当 $\varepsilon\in[0,1]$时，有 $\delta'\le\delta$。因此，失败概率严格降低，从而产生隐私放大效应，具体表述如推论 4.2所示。

**推论 4.2**。 复合机制 $Q^{(\eta)}\circ A$ 满足$(\varepsilon,\delta')$-差分隐私，且 $\delta'<\delta$，因此严格放大了整体失败概率。

基于（Lecuyer 等人，2019年），我们在推论 $Q^{(\eta)}\circ A$ 中推导出复合机制的可认证对抗鲁棒性的显式条件4.3。该条件定义了一个鲁棒性阈值，模型的期望置信度分数必须超过此阈值。值得注意的是，由于推论4.2中形式化的隐私放大效应，复合机制下的鲁棒性阈值（由 $\delta'$参数化）严格低于原始经典机制下的鲁棒性阈值（由 $\delta$参数化）。因此，量子后处理可证明地扩大了能够保证对抗鲁棒性的输入集合。有关对抗鲁棒性的更多细节，请参阅附录 A。

**推论 4.3**。 若输入 $x\in\mathbb{X}$ 对正确类别 $k$ 满足以下条件，则复合机制 $C=Q^{(\eta)}\circ A$ 对抗对抗扰动具有可认证的鲁棒性：

$$\mathbb{E}[[C(x)]_k]>e^{2\varepsilon}\max_{i\neq k}\mathbb{E}[[C(x)]_i]+(1+e^{\varepsilon})\delta'$$

#### 4.2.2. 第二种分析——放大隐私损失

我们研究复合机制 $Q^{(\eta)}\circ A$ 如何同时放大隐私损失 $\varepsilon$ 和失败概率 $\delta$。该结果以定理 4.4 的形式给出，提供了新的 $(\varepsilon',\ \delta')$ 界。证明（详见附录 B）依赖于最初由（高级联合凸性理论，该理论最初由（Balle 等人，2018）提出。核心洞见在于，去极化信道会将最终输出分布转换为原始（无噪声）分布与最大混合态分布的凸组合。这种明确的混合结构使得可以应用联合凸性定理，从而得到关于隐私损失和失败概率的新差分隐私界。

定理 4.4 揭示，放大后的失败概率 $\delta'$取决于正算子值测度的选择。特别是，$\delta'$ 随着 $\varphi=\min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$ 增大而变得更紧。这一洞见导出了推论4.5，其中指出，当所有 POVM 元素 $E_k$ 的迹相等时（即 $\mathrm{Tr}(E_k)=\frac{1}{K}$），$\delta'$ 达到最小值。

相反，$\varepsilon'\le\varepsilon$ 对于所有 $\eta\in[0,1]$，以 $\varepsilon$ 衡量的隐私损失总是会降低。然而，只有当噪声水平 $\delta$ 超过推论 $\delta'\le\delta$中给出的阈值时，$\eta$ 的界才会有所改善（即 4.6）。这一条件表明，必须达到足够的量子噪声水平，才能在两个参数上严格增强隐私保证。

**定理 4.4** （隐私损失放大）。 设 $A=f_{par}\circ f_{cdp}$ 为（$\varepsilon,\delta$）-差分隐私，且 $Q^{(\eta)}=f_{mea}\circ f^{(\eta)}{}_{\text{退极化}}\circ f_{enc}$ 是作用于 $d$维希尔伯特空间中的量子机制

其中 $0 \le \eta \le 1$ 为退极化噪声因子。那么，复合 $Q^{(\eta)} \circ A$ 为 $(\varepsilon', \delta')$-差分隐私，其中 $\varepsilon' = \text{对数}\big(1 + (1-\eta)\ (e^{\varepsilon}-1)\ \big)$ 且 $\delta' = (1-\eta)\ \big(1 - e^{\varepsilon'-\varepsilon}\ (1-\delta) - (e^{\varepsilon} - e^{\varepsilon'})\ \varphi\big)$, 其中 $\varphi = \text{最小值}_k \left( \frac{\mathrm{Tr}(E_k)}{d} \right)$。

**推论 4.5**。设 $\{E_k\}_{k=1}^{K}$ 为 $f_{mea}$中使用的正算子值测度（*POVM*）。那么，定理 $\delta'$ 中的放大后失败概率 *4.4* 在所有 *POVM* 元素的迹相等时（即 迹$[E_k] = \frac{d}{K}$对所有 $k \in \{1, \ldots, K\}$ 均成立）达到最小值。

**推论 4.6**。给定一个满足迹$[E_k] = \frac{d}{K} \forall k$的最优测量，复合机制 $Q^{(\eta)} \circ A$严格改善隐私保证（即，$\varepsilon' \le \varepsilon$且 $\delta' \le \delta$），当

$$\eta \ge 1 - \frac{\delta}{(1-\delta)(1-e^{-\varepsilon}) - (e^{\varepsilon}-1)/K}$$

#### 4.2.3. 第三项分析——推广至其他量子噪声信道

虽然我们前两项分析聚焦于退极化噪声，但隐私放大背后的机制自然也适用于更广泛的一类量子信道。核心观点是，只要量子噪声信道会导致量子曲棍球棒散度发生非平凡收缩，就必然会带来隐私放大。在本小节中，我们将展示这一原理如何将第一项分析推广至两种受到广泛研究的非对称噪声信道：广义振幅阻尼（GAD）和广义退相位（GD）。更详细的讨论见附录 B.4。

**广义振幅阻尼下的放大**。GAD 信道本质上不对称且非幺正。尽管如此，我们证明它会将迹距离收缩至多 $(2\sqrt{\eta} - \eta)$倍，其中 $\eta$ 是 GAD 信道的阻尼强度。将这一收缩代入定理 4.1 的证明框架，可得以下 $(\varepsilon', \delta')$。

**定理 4.7**（广义振幅阻尼噪声下的放大）。令 $A: \mathbb{X} \to \mathcal{P}(\mathbb{Y})$ 为满足 $(\varepsilon, \delta)$-差分隐私的经典机制，其中 $A = f_{par} \circ f_{cdp}$，并令 $Q^{(p,\eta)}: \mathbb{Y} \to \mathcal{P}(\mathbb{Z})$ 为定义在$d$维希尔伯特空间中的量子机制，其定义为 $Q^{(p,\eta)} = f_{mea} \circ f^{(p,\eta)}_{\mathrm{GAD}} \circ f_{enc}$。那么，复合机制 $Q^{(p,\eta)} \circ A$满足 $(\varepsilon', \delta')$-差分隐私，其中

$$\varepsilon' = \varepsilon, \quad \delta' = (2\sqrt{\eta} - \eta)\delta.$$

**赤道编码下的广义退相位**。退相位噪声保留经典布居，但会抑制量子相干性。尽管其最坏情况下的收缩系数为 1，我们证明，对于许多量子机器学习编码方案，包括基于角度的编码器，编码态位于布洛赫球的赤道面上。在这种结构下，所有可区分性都编码在相干项中，而这些相干项会直接受到广义退相位噪声的影响，因此能够产生非平凡的收缩。

**假设 4.8**（所有量子比特上的乘积赤道编码）。对于每个输入 $y \in \mathbb{Y}$，编码器都会制备一个乘积态

$$\rho_y = f_{\mathrm{enc}}(y) = \bigotimes_{j=1}^{n} \rho_y^{(j)},$$

其中，每个单量子比特因子 $\rho_y^{(j)}$ 都是布洛赫球上的赤道态，即

$$\rho_y^{(j)} = \frac{1}{2}\Big(I + \cos\phi_y^{(j)}\, X + \sin\phi_y^{(j)}\, Y\Big),$$

对于某个角度 $\phi_y^{(j)} \in \mathbb{R}$，且没有 $Z$ 分量。

在此假设下，GD 通道会将所有相关相干项收缩 $|1 - 2\eta|$ 倍，其中 $\eta$ 是 GD 通道的退相位参数，从而得到以下隐私保证。

**定理 4.9**。设 $A: \mathbb{X} \to \mathcal{P}(\mathbb{Y})$ 是满足 $(\varepsilon, \delta)$-差分隐私的经典机制，并设

$$Q^{(\eta)} := f_{\mathrm{mea}} \circ f_{\mathrm{GD}}^{(\eta)} \circ f_{\mathrm{enc}}$$

是一个 $n$-量子比特量子机制，其中 $f^{(\eta)}{}_{\mathrm{GD}}$ 是 $n$-量子比特 *GD* 通道，且 $f_{\mathrm{enc}}$ 满足假设 *4.8*。那么复合机制 $Q^{(\eta)} \circ A$ 满足 $(\varepsilon', \delta')$-差分隐私，其中

$$\varepsilon' = \varepsilon, \qquad \delta' = |1 - 2\eta| \cdot \delta.$$

### 4.3. 效用界

最后，我们建立了一个严谨的框架，用于研究效用损失，即机制的含噪版本与无噪版本之间的绝对误差。由于采用基于采样的测量过程，机制的最终输出具有随机性。因此，我们分析其输出期望值之间的差异。期望值代表机制的平均行为，并提供一个确定性量，可用于衡量效用损失。

形式化地，我们定义期望测量函数$f_{\mathrm{exp}}$： $\mathcal{H} \to \mathbb{R}$如下：

$$f_{\mathrm{exp}}(\rho) = \sum_k k\,\mathrm{Tr}[E_k \rho] = \mathrm{Tr}[E_{\mathrm{exp}}\rho]$$

其中， $E_{\mathrm{exp}} = \sum_k k E_k$ 是期望值可观测量。

利用该函数，我们定义确定性期望机制。**完整机制**，包括经典噪声和

量子噪声，是 $\mathcal{M}_{\text{full}}(x) = (f_{\text{exp}} \circ f^{(\eta)}_{\text{dep}} \circ f_{\text{enc}} \circ f_{\text{par}} \circ f_{\text{cdp}})(x)$。另一方面，**无噪声机制（clean）**是 $\mathcal{M}_{\text{clean}}(x) = (f_{\text{exp}} \circ f_{\text{enc}} \circ f_{\text{par}})(x)$。总效用损失是两者期望输出之间的最坏情况绝对误差：

$$\text{Error} = \sup_{x \in \mathbb{X}} |\mathcal{M}_{\text{full}}(x) - \mathcal{M}_{\text{clean}}(x)|$$

**定理 4.10** (效用界)。 设作用于维度为$d_X = \dim(\mathbb{X})$的输入空间 $\mathbb{X}$ 上的经典噪声为$\kappa \sim \mathcal{N}(0, \sigma^2 I)$。对于任意给定的失败概率 $p > 0$，效用损失的概率上界为：

$$\Pr\left(\mathit{Error} \leq L_\infty \cdot \sigma \sqrt{2 \ln \frac{2d_X}{p}} + 2\eta \|E_{exp}\|_{op}\right) \geq 1 - p$$

*where* $L_\infty = 2(1-\eta)\|E_{exp}\|_{op}\|W\|_\infty \left(\sum_j \|H_j\|_{op}\right)$.

定理 4.10 给出了一个效用界，用于量化隐私与性能之间的权衡。证明（详见附录 B）使用了一种**中间机制（半机制）**，该机制仅包含量子噪声，即 $\mathcal{M}_{\text{half}}(x) = (f_{\text{exp}} \circ f^{(\eta)}_{\text{依赖}} \circ f_{\text{enc}} \circ f_{\text{par}})(x)$。具体而言，首先，我们对量子噪声（$|\mathcal{M}_{\text{half}} - \mathcal{M}_{\text{clean}}|$）引入的误差进行界定，并证明该误差与量子噪声水平$\eta$成正比。其次，我们通过为纯量子机制建立利普希茨常数 $L_\infty$，对经典噪声产生的误差进行界定。由于经典噪声无界，最终的保证是一个高概率陈述，它将效用损失与经典（$\sigma$）和量子（$\eta$）噪声水平联系起来。

## 5. 实验评估

在本节中，我们通过实证评估 HYPER-Q，重点关注对抗鲁棒性，这是推论 4.3 中差分隐私保证的直接结果。首先，我们对退极化噪声参数 $\eta$ 进行敏感性分析，以确定量子失真与隐私放大之间的最佳平衡。随后，我们表明，在给定固定隐私预算 $(\varepsilon', \delta')$ 的情况下，HYPER-Q 对于量子机器学习模型而言，比现有差分隐私机制更高效。

**实现细节**。 我们实现了一个旨在融入 HYPER-Q 的量子机器学习模型。在量子机器学习架构中，变分量子电路使用 5-10 个量子比特，由三个交替排列的单量子比特 RX 旋转层和纠缠层组成，分别对应编码函数 $f_{\text{enc}}$。纠缠通过环形排列的 CNOT 门实现，随后进行计算基投影测量，并通过最后的全连接层生成预测。这种架构在以往的量子机器学习和量子差分隐私研究中得到广泛采用（Havlíˇcek 等人，2019年；Du 等人，2021a；Watkins 等人，2023年；Long 等人，2025年）。实现采用 PennyLane 库（Bergholm 等人，2022年），量子电路在模拟器上运行，这是对量子应用进行原型设计和评估的标准做法（Cicero 等人，2025年）。为确保差分隐私，直接向输入添加高斯噪声，并在量子电路中将退极化噪声作为一层施加。具体而言，给定目标隐私预算 $(\varepsilon', \delta')$，退极化噪声水平 $\eta$ 固定不变，而高斯噪声水平 $\sigma^2$ 则根据定理计算得出4.1。更多细节及论证见附录C。

**数据集与基线**。 我们在三个标准图像分类数据集上评估所提出的方法：MNIST（Lecunetal., 1998）、FashionMNIST（Xiao 等人，2017）和 USPS（Hull, 2002）。我们将 HYPER-Q 与经典差分隐私机制进行比较，包括基础高斯、解析高斯（Balle& Wang，2018）和 DP-SGD（Abadi 等人，2016年；Watkins 等人，2023年）。需要指出的是，前两种机制在输入层添加噪声，而 DP-SGD 则在量子机器学习模型的梯度层注入噪声。每项实验报告 10 次运行的平均准确率。更多详情见附录 D。

**对抗鲁棒性设置**。我们采用经过认证的防御框架（Lecuyer 等人，2019年），使用由差分隐私预算校准的噪声层训练模型$(\varepsilon', \delta')$，并设定构造攻击界 $L_{\text{cons}}$。随后，我们通过测量模型对 FGSM（Goodfellow 等人，2015年）和 PGD（Madry 等人，2018攻击的准确率来评估鲁棒性，这些攻击的强度由经验攻击界 $L_{\text{attk}}$定义。更多详情见附录E。

### 5.1. $\eta$的敏感性分析

为了表征去极化参数对模型效用的影响，我们评估了不同退极化噪声水平 $\eta \in \{0.05, 0.1, \ldots, 0.4\}$ 下的模型准确率。对于每个噪声水平 $\eta$，我们测量了模型在不同强度 $L_{\text{attk}} \in \{0, 0.01, \ldots, 0.05\}$ 的 FGSM攻击下的鲁棒性。我们在三个数据集（MNIST、Fashion-MNIST 和 USPS）上，针对四种不同的差分隐私保证 $\varepsilon' \in \{0.25, 0.5, 0.75, 1\}$ 开展了这些实验。

如图 2所示， $\eta$ 与模型准确率之间的关系，在所有评估设置中都呈现出极为一致的单峰结构。准确率曲线遵循共同的模式。具体而言，随着 $\eta$ 升高，效用最初会增加。这一趋势源于量子噪声能够减轻对具有破坏性的经典高斯噪声的需求。随后，当$\eta = 0.1$ 为 $\varepsilon' \leq 0.5$且 $\eta = 0.05$ 为 $\varepsilon' > 0.5$时，效用达到峰值。最后，随着量子失真盖过底层数据信号，效用会下降。我们的敏感性分析揭示了稳定的趋势，为退极化噪声校准提供了可靠的经验准则。首先，在所有隐私预算和攻击强度下， $\eta$ 的最优值始终落在 $[0.05, 0.15]$ 这一较窄区间内。此外，较宽松的

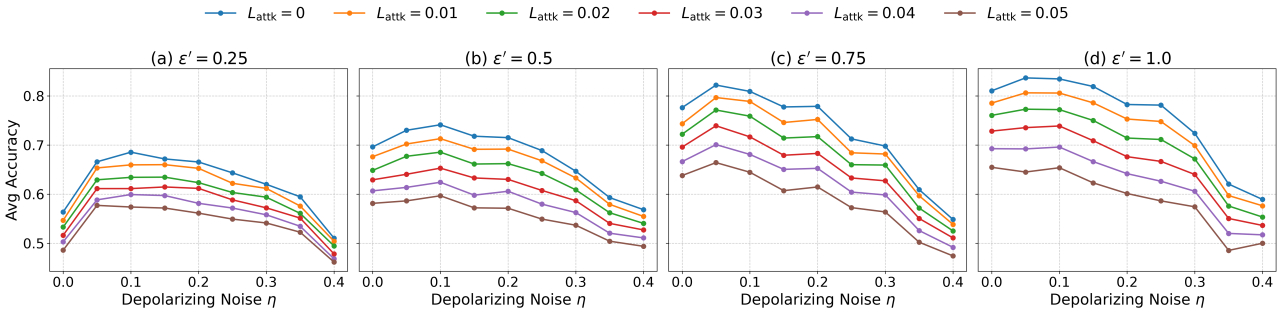



图 2. 子图 (a)–(d) 展示了在隐私预算为 $\varepsilon' \in \{0.25, 0.5, 0.75, 1.0\}$ 时，MNIST、FashionMNIST 和 USPS 数据集上量子噪声与模型准确率之间的权衡。每条曲线表示不同强度的 FGSM 攻击（$L_{\text{attk}}$）下的平均准确率。

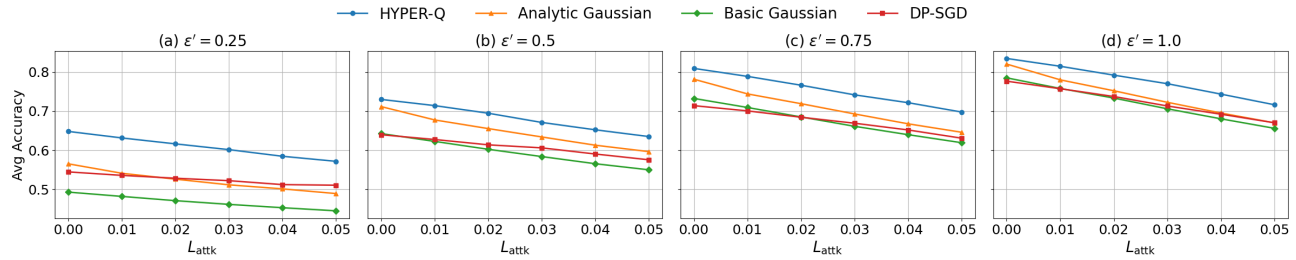



图 3. 在 MNIST、FashionMNIST 和 USPS 数据集上，加噪机制在 FGSM 和 PGD 攻击下的平均准确率。准确率 y 在所有 $L_{\text{cons}}$ 设置中取平均值，每个设置对应 $(L_{\text{attk}}, \varepsilon')$。使用 $\eta = 0.1$ 和 $\delta' = 1 \times 10^{-5}$ 对 HYPER-Q 进行评估。

隐私预算通常需要较小的 $\eta$ 值才能获得最佳结果。更多详情见附录F.1。

### 5.2. HYPER-Q 的鲁棒性分析

在本实验中，我们展示了在相同隐私预算下，HYPER-Q 在量子机器学习中比经典机制更有效地保留对抗鲁棒性。我们评估了 HYPER-Q 在退极化噪声 $\eta = 0.1$ 下的对抗鲁棒性，该噪声是根据我们的敏感性分析选取的。我们将其性能与基础高斯、解析高斯和 DP-SGD 机制进行比较。为确保比较公平，我们保证所有方法都在相同隐私预算下评估，并应用于相同的量子机器学习模型。

图 3 展示了 MNIST、FashionMNIST 和 USPS 数据集在 FGSM 与 PGD 攻击下的平均准确率，对应四种不同的隐私预算 $\varepsilon' \in \{0.25, 0.5, 0.75, 1\}$。我们观察到，HYPER-Q 在无攻击（$L_{\text{attk}} = 0$）与有攻击（$L_{\text{attk}} > 0$）两种情况下均持续优于所有基线方法。随着 $\varepsilon'$增大，性能差距变得更加显著。具体而言，HYPER-Q 平均以 16.54%、5.37%、6.44%和5.20% 的准确率优势超越次优方法解析高斯机制，对应四个相应的 $\varepsilon'$ 取值。这表明，用量子噪声合理替代一定量的经典噪声可以提升对抗准确率。更多细节见附录F.2。

为进行全面评估，我们请读者参阅附录中的其他实验研究，包括：（i）与经典机器学习模型的对比基准数据集，其中配备HYPER-Q 的量子机器学习模型与受经典差分隐私机制保护的经典模型进行比较（附录F.3）；（ii）维度可扩展性分析（附录F.4）；（iii）对定理4.10中效用界紧致性的实证验证（附录F.5）；以及（iv）在 CIFAR-10 数据集上的评估（附录F.6）。

## 6. 结论

在本研究中，我们提出了 HYPER-Q，一种面向量子系统的混合隐私保护机制。通过在三个真实世界数据集上进行广泛的实验分析，并对其施加 FGSM 和 PGD 攻击，我们证明了量子噪声与经典噪声的结合既稳健又可扩展，同时显著提升了隐私保护和效用。经典组件确保了训练的稳定性和可解释性，而量子噪声则引入了天然随机性，在不显著降低效用的情况下增强隐私保护。随着量子硬件日益成熟，我们预计 HYPER-Q 这类框架对于塑造隐私保护机器学习的未来将至关重要。未来工作的一个核心方向是研究混合差分隐私机制在实际量子硬件上运行的更大型变分电路中的表现。

## 影响声明

本文介绍的研究旨在推动机器学习领域的发展。这项研究可能会对社会产生多方面的影响。

，其中没有任何内容是我们认为必须在此特别强调的。

## 参考文献


Abadi, M.、Chu, A.、Goodfellow, I.、McMahan, H. B.、Mironov, I.、Talwar, K. 和 Zhang, L. 采用差分隐私的深度学习。载于 *2016*年*ACM SIGSAC*计算机与通信安全会议论文集，第308–318页，2016年。Bai, Y.-R.、Tao, Y.-H.、Wu, S.-H.、Zhang, H. 和 Fei, S.- M. 噪声信道下的量子差分隐私。*PhysicaScripta*，99(3):035119，2024年。Balle, B. 和 Wang, Y.-X. 改进差分隐私的高斯机制：解析校准与最优去噪。载于 Dy, J. 和 Krause, A.（编），第*35*届机器学习国际会议论文集，第80卷，机器学习研究会议录，第394–403页。PMLR，2018年7月10日至15日。网址：
h
t
tps://proceedings.mlr.press/v80/balle18a.html。
Balle, B.、Barthe, G. 和 Gaboardi, M. 子采样带来的隐私放大：通过耦合与散度进行紧致分析。载于第*32*届神经信息处理系统国际会议论文集，NIPS'18，第6280–6290页，美国纽约州雷德胡克，2018年。Curran Associates Inc. Balle, B.、Barthe, G.、Gaboardi, M. 和 Geumlek, J. 通过混合与扩散机制实现隐私放大。载于神经信息处理系统，2019年。网址：
h
t
t
p
s://
api.semanticscholar.org/CorpusID:168170121。
Balle, B.、Bell, J.、Gasc´on, A. 和 Nissim, K. 洗牌模型的隐私毯。载于密码学进展 – *CRYPTO 2019*：第*39*届年度密码学会议，美国加利福尼亚州圣巴巴拉，*2019*年*8*月*18*日至*22*日，论文集，第二部分，第638–667页，柏林、海德堡，2019年。施普林格出版社。ISBN 978-3-030-26950-0。doi：10.1007/978-3-030-26951-7 22。网址：
https://doi.org/10.1007/978-3-030-26951-7_22。
Berberich, J.、Fink, D. 和 Holm, C. 量子算法对相干控制误差的鲁棒性。*Phys.Rev.A*，109:012417，2024年1月。doi：10.1103/PhysRevA.109. 012417。网址：
h
t -
t
p
s://
link.aps.org/doi/10.1103/PhysRevA.109.012417。
Bergholm, V.、Izaac, J.、Schuld, M.、Gogolin, C.、Ahmed, S.、Ajith, V.、Alam, M. S.、Alonso-Linaje, G.、Akash-Narayanan, B.、Asadi, A.、Arrazola, J. M.、Azad, U.、Banning, S.、Blank, C.、Bromley, T. R.、Cordier, B. A.、Ceroni, J.、Delgado, A.、Matteo, O. D.、Dusko, A.、Garg, T.、Guala, D.、Hayes, A.、Hill, R.、Ijaz, A.、Isacsson, T.、Ittah, D.、Jahangiri, S.、Jain, P.、Jiang, E.、Khandelwal, A.、Kottmann, K.、Lang, R. A.、Lee, C.、Loke, T.、Lowe, A.、McKiernan, K.、Meyer, J. J.、Monta˜nez-Barrera, J. A.、Moyard, R.、Niu, Z.、O'Riordan, L. J.、Oud, S.、Panigrahi, A.、Park, C.-Y.、Polatajko, D.、Quesada, N.、Roberts, C.、S´a, N.、Schoch, I.、Shi, B.、Shu, S.、Sim, S.、Singh, A.、Strandberg, I.、Soni, J.、Sz´ava, A.、Thabet, S.、Vargas- Hern´andez, R. A.、Vincent, T.、Vitucci, N.、Weber, M.、Wierichs, D.、Wiersema, R.、Willmann, M.、Wong, V.、Zhang, S.与 Killoran, N. PennyLane：混合量子-经典计算的自动微分，2022年。网址 https://arxiv.org/abs/1811.04968。Bun, M.、Nissim, K.、Stemmer, U.与Vadhan, S. P. 差分隐私发布与阈值函数的学习。*2015*年*IEEE*第*56*届计算机科学基础年度研讨会，第634–649页，2015年。网址
https://api.semanticscholar.org/CorpusID:3338015。
Cheu, A.、Smith, A. D.、Ullman, J.、Zeber, D.与Zhilyaev, M. 通过洗牌实现分布式差分隐私。国际密码学研究协会密码学电子预印本档案，2019:245，2018年。网址
h
ttps://api.semanticscholar.org/CorpusID:54464170。
Cicero, A.、Maleki, M. A.、Azhar, M. W.、Kockum, A. F.与 Trancoso, P. 量子计算机模拟：综述与加速机遇。*ACM*量子计算汇刊，2025年9月。doi: 10.1145/3762672。网址
https://doi.org/10.1145/3762672。已录用待刊。Cohen, J.、Rosenfeld, E.与Kolter, Z. 通过随机平滑实现经认证的对抗鲁棒性。见Chaudhuri, K.与Salakhutdinov, R.（编），第*36*届国际机器学习会议论文集，机器学习研究论文集第97卷，第1310–1320页。PMLR，2019年6月9–15日。网址
https://proceedings.mlr.press/v97/cohen19c.html。
Cullen, A. C.、Montague, P.、Liu, S.、Erfani, S. M.与 Rubinstein, B. I. P. 就是Simplex！分解度量以提升经认证的鲁棒性。载于*2024*年*IEEE*安全与隐私研讨会（*SP*），第2886–2900页，美国加利福尼亚州洛斯阿拉米托斯，2024年5月。IEEE计算机学会。doi: 10.1109/SP54263.2024.00065。网址
h
t
t
p
s://
d
o
i.
i
e
eecomputersociety.org/10.1109/SP54263.2024.00065。

Dong, J.、Roth, A.与Su, W. J. 高斯差分隐私。皇家统计学会期刊*B*辑：统计方法论，84(1):3–37，2022年。

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S. 等。一张图像胜过 16×16 个词：用于大规模图像识别的Transformer。载于国际学习表征会议，2021。

Du, Y., Hsieh, M.-H., Liu, T., Tao, D. 和 Liu, N. 量子噪声保护量子分类器免受对手攻击。*Phys. Rev. Res.*，3:023153，2021年5月。

Du, Y., Hsieh, M.-H., Liu, T., Tao, D. 和 Liu, N. 量子噪声保护量子分类器免受对手攻击。 *Phys. Rev.Res.*，3:023153，2021年5月。doi: 10.1103/PhysRevResearch.3.023153。URL
h
t
tps://link.aps.org/doi/10.1103/PhysRevResearch.3.023153。

Dwork, C. 差分隐私。载于国际自动机、语言与编程研讨会，2006年。网址：
https://api.semanticscholar.org/CorpusID:2565493。

Dwork, C. 和 Roth, A. 差分隐私的算法基础。计算机科学理论与实践基金会趋势，9(3–4):211–407，2014年8月。ISSN 1551-305X。doi: 10.1561/0400000042。网址：
https://doi.org/10.1561/0400000042。

Dwork, C.、McSherry, F.、Nissim, K. 和 Smith, A. 为隐私数据分析中的敏感度校准噪声。载于密码学理论会议，第265–284页。施普林格，2006年。

Erlingsson, U.、Feldman, V.、Mironov, I.、Raghunathan, A.、Talwar, K. 和 Thakurta, A. 通过洗牌实现放大：借助匿名性从本地差分隐私到中心化差分隐私。载于第三十届 *ACM-SIAM* 离散算法年度研讨会论文集，SODA ’19，第2468–2479页，美国，2019年。工业与应用数学学会。

Feldman, V., Mironov, I., Talwar, K., and Thakurta, A. 通过迭代实现隐私放大。*2018*年 *IEEE* 第 *59* 届计算机科学基础年度研讨会（*FOCS*），第 521–532 页，2018年。网址
h
ttps://api.semanticscholar.org/CorpusID:52055752。

Fisher, K. A. G., Broadbent, A., Shalm, L. K., Yan, Z., Lavoie, J., Prevedel, R., Jennewein, T. 和 Resch, K. J. 基于加密数据的量子计算。《自然·通讯》，5(1):3074，2014年1月。ISSN 2041-1723。doi: 10.1038/ncomms4074。URLhttps://doi.org/10.1038/ncomms4074。

Geng, Q. 和 Viswanath, P. 差分隐私中的最优加噪机制。*IEEE* 信息论汇刊，62:925–951，2012年。URLhttps://api.semanticscholar.org/CorpusID:8989005。

Ghazi, B., Guzm´an, C., Kamath, P., Kumar, R., and Manu- rangsi, P. 使用稀疏梯度进行差分隐私优化。载于第 *38* 届神经信息处理系统国际会议论文集，NIPS ’24，美国纽约州雷德胡克，2025年。Curran Associates Inc. ISBN 9798331314385。

Goodfellow, I. J., Shlens, J., and Szegedy, C. 解释并利用对抗样本。国际学习表征会议，2015年。

Hatakeyama-Sato,K., Igarashi,Y.,Kashikawa, T.,Kimura, K., 和 Oyaizu, K. 量子电路学习作为预测实验化学性质的潜在算法。 *DigitalDiscovery*, 2:165–176, 2023年. doi: 10.1039/D2DD00090C. URLhttp://dx.doi.org/10.1039/D2DD00090C.

Havl´ıˇcek,V.,C´orcoles, A. D., Temme, K., Harrow, A. W., Kandala, A., Chow, J. M., 和 Gambetta, J. M. 基于量子增强特征空间的监督学习。《自然》, 567(7747):209–212, 2019年3月. ISSN 1476- 4687. doi: 10.1038/s41586-019-0980-2. URLhttps://doi.org/10.1038/s41586-019-0980-2.

He, K., Zhang, X., Ren, S., 和 Sun, J. 用于图像识别的深度残差学习。载于*IEEE* 计算机视觉与模式识别会议论文集，第 770–778 页，2016年。

Hirche, C. 量子Doeblin系数：收缩系数的一个简单上界。载于*2024*年 *IEEE* 信息论国际研讨会（*ISIT*），第 557–562 页，2024年。doi：10.1109/ISIT57864.2024.10619667。

Hirche, C., Rouz´e, C., 和 Franc¸a, D. S. 量子差分隐私：信息论视角。*IEEE* 信息论汇刊，69(9)：5771–5787，2023年。doi：10.1109/TIT.2023.3272904。

Hu, Z., Wolle, R., Tian, M., Guan, Q., Humble, T., 和 Jiang, W. 通过高效的原位校准，在不稳定设备上实现一致的高保真量子学习。载于 Muller, H., Alexev, Y., Delgado, A., 和 Byrd, G.（编），论文集——*2023 IEEE*量子计算与工程国际会议，*QCE 2023*，论文集——2023 IEEE量子计算与工程国际会议，QCE 2023，第848–858页。电气与电子工程师协会，2023年。doi: 10.1109/QCE57702.2023.00099。出版版权：© 2023 IEEE；第4届 IEEE 国际量子计算与工程会议

，QCE 2023；会议日期：2023年9月17日至2023年9月22日。

Hull, J. J. 手写文本识别研究数据库。*IEEE*模式分析与机器智能汇刊，16(5)：550–554，2022年。

Hur, T.、Kim, L. 和 Park, D. K. 用于经典数据分类的量子卷积神经网络。量子机器智能，4(1)：3，2022年2月。ISSN 2524- 4914。doi：10.1007/s42484-021-00061-x。
U
RLhttps://doi.org/10.1007/s42484-021-00061-x。

Ji, T. 和 Li, P. 少即是多：重新审视用于差分隐私的高斯机制。载于 第*33*届*USENIX*安全专题会议论文集，SEC '24，美国，2024年。USENIX协会。ISBN 978-1-939133- 44-1。

Koga, T., Meehan, C. 和 Chaudhuri, K. 时间域中的子采样隐私放大。载于 Camps-Valls, G., Ruiz, F. J. R. 和 Valera, I.（编），第*25*届国际人工智能与统计会议论文集，第151卷，机器学习研究会议录，第4055–4069页。PMLR，2022年3月28–30日。网址：
h
t
t
p
s://
proceedings.mlr.press/v151/koga22a.html。
Lecun, Y., Bottou, L., Bengio, Y. 和 Haffner, P. 应用于文档识别的基于梯度的学习。*IEEE*论文集，86(11):2278–2324，1998年。doi: 10.1109/5.726791。Lecuyer, M., Atlidakis, V., Geambasu, R., Hsu, D. 和 Jana, S. 通过差分隐私实现对抗样本的经认证鲁棒性。载于*2019*年 *IEEE*安全与隐私研讨会（*SP*），第656–672页，2019年。doi: 10.1109/SP. 2019.00044。Long, C., Huang, M., Ye, X., Futamura, Y. 和 Sakurai, T. 混合量子-经典-量子卷积神经网络。《科学报告》，15(1): 31780，2025年8月。ISSN 2045-2322。doi: 10.1038/ s41598-025-13417-1。网址：
https://doi.org/10.1038/s41598-025-13417-1。
Madry, A., Makelov, A., Schmidt, L., Tsipras, D. 和 Vladu, A. 面向能够抵御对抗攻击的深度学习模型。载于国际学习表征会议，2018年。Melo, A., Earnest-Noble, N. 和 Tacchino, F. 高效脉冲量子机器学习。*Quantum*，7:1130，2023年10月。ISSN 2521-327X。doi: 10.22331/q-2023-10-09-1130。网址：
https://doi.org/10.22331/q-2023-10-09-1130。

Nielsen, M. A. 和 Chuang, I. L.《量子计算与量子信息：十周年纪念版》。剑桥大学出版社，2010年。

P´erez-Salinas, A.、Cervera-Lierta, A.、Gil-Fuster, E.与 Latorre, J. I. 用于通用量子分类器的数据重上传。 *Quantum*，4:226，2020年2月。ISSN 2521- 327X。doi: 10.22331/q-2020-02-06-226。网址
https://doi.org/10.22331/q-2020-02-06-226。

Phan, H.、Thai, M. T.、Hu, H.、Jin, R.、Sun, T.与Dou, D. 对抗学习中具有经认证的鲁棒性的可扩展差分隐私。载于国际机器学习会议，2019年。网址
h
t
tps://api.semanticscholar.org/CorpusID:219558337。

Schuld, M. 与 Killoran, N. 特征希尔伯特空间中的量子机器学习。《物理评论快报》，122: 040504，2019年2月。doi: 10.1103/PhysRevLett.122. 040504。网址
h
t
t
p
s://
l
i
nk.aps.org/doi/10.1103/PhysRevLett.122.040504。
Song, B.、Pokhrel, S. R.、Vasilakos, A. V.、Zhu, T. 与 Li, G. 迈向混合量子差分隐私。 *IEEE*通信领域精选期刊，2025年。Wang, Y.-X.、Balle, B. 与 Kasiviswanathan, S. P. 子采样 r´enyi 差分隐私与解析矩会计器。载于 第*22*届国际人工智能与统计会议，第1226–1235页。PMLR，2019年。Watkins, W. M.、Chen, S. Y.-C. 与 Yoo, S. 带差分隐私的量子机器学习。《科学报告》，13(1):2453，2023年。Xiao, H.、Rasul, K. 与 Vollgraf, R. Fashion-MNIST：一种用于基准评测机器学习算法的新型图像数据集。 *arXiv*预印本 *arXiv:1708.07747*，2017。Xiao, X.、Yao, Y.、Zhou, L.-M. 与 Wang, X. 非对称克隆机中量子费 Fisher 信息的分布。《科学报告》，4(1):7361，2014年12月。ISSN 2045-2322。doi: 10.1038/srep07361。网址https://doi.org/10.1038/srep07361。Ye, J. 与 Shokri, R. 差分隐私学习需要隐藏状态（或快得多的收敛）。载于 Koyejo, S.、Mohamed, S.、Agarwal, A.、Belgrave, D.、Cho, K. 与 Oh, A.（编）， 神经信息处理系统进展，第35卷，第703–715页。Curran Associates, Inc.，2022年。Yuan, X.、Ni, W.、Ding, M.、Wei, K.、Li, J. 与 Poor, H. V. 用于在联邦学习中平衡隐私与效用的振幅变化扰动。《*IEEE*信息取证与安全汇刊》，18:1884–1897，2023年。doi: 10.1109/TIFS.2023.3258255。

Zhou, L. 与 Ying, M. 量子计算中的差分隐私。载于 *2017 IEEE第30届计算机安全基础研讨会（CSF）*，第249–262页，2017。doi: 10.1109/CSF.2017.23。

# 附录

## A. 其他背景

### A.1. 量子神经网络

量子神经网络（量子神经网络）是一类量子机器学习模型，利用参数化量子电路从经典或量子数据中进行学习。在本文中，我们重点研究针对经典输入设计的量子神经网络。在监督学习环境下，量子神经网络旨在通过在数据集 $S = \{(x_i,\ y_i)$ 上进行训练，近似未知函数 $K$： $\mathbb{X} \to \mathbb{Y}$，其中每个 $x_i \in \mathbb{R}^d$ 都是一个输入数据向量，而 $y_i$ 是与之对应的标签。

量子神经网络模型使用参数化量子电路处理数据。典型量子神经网络的工作流程包括：

- **数据编码：** 经典数据通过参数化的“编码器”电路映射到量子比特的量子态中。这一步至关重要，因为编码器可以经过训练，以找到强大的数据表示，并引入纠缠等量子特征，从而提升模型的能力。

- **模型电路：** 一系列参数化量子门（类似于经典神经网络的各层）对编码后的量子态进行处理。

- **测量：** 对最终状态进行测量，以提取经典输出，作为模型的预测。

训练 QNN 是一个量子-经典混合过程。量子计算机执行电路并进行测量。随后，经典计算机计算损失函数（例如均方误差），以量化预测与真实标签之间的误差。给定可微损失函数 $f(\cdot)$，目标是最小化：

$$\mathcal{L}(\theta) = \sum_{i=1}^{N} f(\ell_i(\theta; y_i), y_i).$$

最后，经典计算机使用基于梯度的优化来更新电路参数 $\theta$。重复这一过程，直至模型收敛。目标是找到使损失最小化的最优参数 $\theta^*$：

$$\theta^* = \arg\min_{\theta} L(\theta).$$

### A.2. 对抗鲁棒性

如果一个模型即使其输入受到恶意扰动而发生轻微改变，仍能持续做出正确预测，则称其具有对抗鲁棒性。这些经过改变的输入称为对抗样本。形式化地，我们定义模型$f$： $\mathbb{X} \to \mathbb{Y}$，它将空间 $\mathbb{X}$ 中的输入映射为标签 $y = \{y_1, y_2, \ldots, y_k\} \in \mathbb{Y}$上的输出分布。如果向 $x$添加小扰动 $\alpha$ 后，模型$f$ 对输入 $x$ 的预测保持不变，则认为该模型具有对抗鲁棒性。这可以表述为：

$$\max_{i\in[1,k]} [f(x)]_i = \max_{i\in[1,k]} [f(x+\alpha)]_i, \quad \forall \alpha \in B_p(L),$$

其中， $B_p$（$L_{\text{cons}}$）表示半径为 $L_{\text{cons}}$的 $p$-范数球，将扰动大小限制为 $\|\alpha\|_p \leq L_{\text{cons}}$。我们也将$L_{\text{cons}}$ 称为构造界。

近年来，差分隐私（差分隐私）已成为增强模型鲁棒性的一种有前景的方法。差分隐私最初用于保护统计数据库中的个人数据，确保当添加或移除某个个体的数据时，算法的输出不会发生显著变化。这通常通过向算法的计算过程中注入经过精心校准的随机性来实现。预测稳定性这一特性构成了差分隐私与对抗鲁棒性之间联系的基础，正如（Lecuyer 等人，2019年）中的研究所探讨的那样。根据设计，使用差分隐私噪声训练的模型对微小输入扰动的敏感性天然较低，从而增强了其抵御对抗攻击的能力。

形式上，给定模型 $f$，它是 $(\varepsilon,\ \delta)$ 差分隐私的（在 $p$-范数度量下），则可保证其对对抗扰动 $\alpha$ （大小为 $\|\alpha\|_p \leq 1$）具有鲁棒性，如果以下条件成立（Lecuyer 等人，2019年）：

$$E([f(x)]_k) > e^{2\varepsilon} \max_{i:i\neq k} E([f(x)]_i) + (1+e^{\varepsilon})\delta, \exists k \in K,$$

其中 $E([f(x)]_k)$ 是正确标签 $k$ 的期望置信度， $E([f(x)]_i)$ 是其他标签的期望置信度。

该条件可证明，任何满足该不等式的输入，都能在规定的扰动大小范围内抵御对抗攻击。更强的差分隐私保证（即 $\varepsilon$ 和 $\delta$的值更小）会扩大具有这种鲁棒性的输入集合。在本研究中，我们旨在探究量子噪声如何增强差分隐私保证，从而显著提升模型的整体鲁棒性。

## B. 技术证明

本节将详细证明主要理论结果。

### B.1. 定理 4.1 的证明与支撑引理

我们研究在固定隐私损失参数 $\varepsilon$ 的假设下，失败概率如何在量子后处理中放大。具体而言，我们旨在对以下量给出上界：

$$\sup_{x,x'} \mathrm{D}_{e^{\varepsilon}} \left( Q^{(\eta)} \circ A(x) \,\|\, Q^{(\eta)} \circ A(x') \right) \tag{1}$$

**引理 B.1.** 设 $\mu$ 和 $\nu$ 是满足 $\mathrm{D}_{e^{\varepsilon}}(\mu\|\nu) \leq \delta$的概率分布，并定义 $\theta = \mathrm{D}_{e^{\varepsilon}}(\mu\|\nu)$。那么，存在分布 $\mu'$、 $\nu'$和 $\omega$，以及参数 $\tilde{\varepsilon} := \log\left(1 + \frac{e^{\varepsilon}-1}{\theta}\right)$，使得：

$$\mu = (1-\theta)\omega + \theta\mu', \quad \nu = \frac{1-\theta}{e^{\varepsilon}}\omega + \left(1 - \frac{1-\theta}{e^{\varepsilon}}\right)\nu',$$

具有不相交的分布： $\mu' \perp \nu'$。那么，以下界成立：

$$D_{e^{\varepsilon}}(\mu\|\nu) = \theta \cdot D_{e^{\tilde{\varepsilon}}}(\mu'\|\nu')$$

证明。 研究见（Balle 等人，2019a） □

将 $A$ 的输出分布记为 $\mu = A(x)$ 和 $\nu = A(x')$，其中 $\mu, \nu \in \mathcal{P}(\mathbb{Y})$。引理 B.1 最初由（Balle 等人，2019a）研究，它根据两个分布 $\mu$ 和 $\nu$ 之间的散度$\theta = \mathrm{D}_{e^{\varepsilon}}(\mu\|\nu)$，建立了它们的分解。具体而言， $\mu$ 被分解为重叠分量 $\omega$ 与残差分量$\mu'$的混合，而 $\nu$ 也类似地被分解为 $\omega$ 与残差分量 $\nu'$。共享分量 $\omega$ 通过密度$p_\omega = \min(p_\mu, e^{\varepsilon} p_\nu)$ $1-\theta$ 定义。剩余分布 $\mu'$ 和 $\nu'$ 分别对应于 $\mu$ 和 $\nu$的不重叠部分，并且可以证明它们具有不相交支撑（即 $\mu' \perp \nu'$）。引理 B.1 还给出了 $\mu$ 与 $\nu$ 之间的散度关于其各自分量 $\mu'$ 和 $\nu'$之间散度的变换，具体为 $\mathrm{D}_{e^{\varepsilon}}(\mu\|\nu) = \theta \cdot \mathrm{D}_{e^{\tilde{\varepsilon}}}(\mu'\|\nu')$。由于量子过程 $Q$ 是一个线性映射（Nielsen与Chuang，2010），它保持输入分布的凸组合。因此，我们得到

$$\mathrm{D}_{e^{\varepsilon}}(\mu Q\|\nu Q) = \theta \cdot \mathrm{D}_{e^{\tilde{\varepsilon}}}(\mu' Q\|\nu' Q) \tag{2}$$

此外， $\mu'$ 与 $\nu'$ 的正交性在分析后处理机制的收缩行为方面发挥着至关重要的作用，这一点将在下面的引理中得到证明。

**引理 B.2**。 给定一个后处理机制 $Q$，我们有：

$$\sup_{\mu\perp\nu} D_{\varepsilon}(\mu Q\|\nu Q) \leq \sup_{y\neq y'} D_{\varepsilon}(Q(y)\|Q(y'))$$

证明。 研究见（Balle 等人，2019年） □

引理B.2 确立了在应用后处理机制后两个正交分布之间散度的上界。令$\tau_y$ 表示位于 $y$处的点质量分布，即若 $\tilde{y} = y$，则$\tau_y(\tilde{y}) = 1$，否则$\tau_y(\tilde{y}) = 0$。于是，$\mu = \sum_{y\in\text{supp}(\mu)} \mu(y)\tau_y$，对 $\nu$同样成立。由$\mathrm{D}_{e^\varepsilon}$ 的凸性以及 $Q$的线性性，我们有：

$$\mathrm{D}_{e^\varepsilon}(\mu Q\|\nu Q) \le \sup_{y\neq y'} \mathrm{D}_{e^\varepsilon}(\tau_y Q\|\tau_{y'} Q) \le \sup_{y\neq y'} \mathrm{D}_{e^\varepsilon}(Q(y)\|Q(y'))$$

综合而言，引理B.1与B.2 阐明了 $A$ 的输出之间的散度在后处理下如何变换。接下来仍需分析仅由 $Q^{(\eta)}$引起的散度，从而使我们能够专注于给出上界：

$$\sup_{y\neq y'} \mathrm{D}_{e^\varepsilon}(Q^{(\eta)}(y)\|Q^{(\eta)}(y')) \tag{3}$$

**引理 B.3.** 给定满足 $\sum_i E_i = I$的测量 $E = \{E_i\}$，以及两个量子态 $\rho$ 与 $\rho'$，所得概率分布的经典曲棍球棒散度小于或等于这些量子态之间的量子曲棍球棒散度。

$$\mathrm{D}_\alpha(P \parallel P') \le D_\alpha^{(q)}(\rho \parallel \rho')$$

证明。 量子曲棍球棒散度定义如下：

$$D_\alpha^{(q)}(\rho \parallel \rho') = \mathrm{Tr}\big[(\rho - \alpha\rho')_+\big],$$

其中，$A_+$ 表示厄米算子 $A$的正部。定义算子 $A = \rho - \alpha\rho'$。
对 $\rho$ 和 $\rho'$ 应用测量 $E$，得到的概率分布元素为：

$$P(i) = \mathrm{Tr}(E_i\rho), \quad P'(i) = \mathrm{Tr}(E_i\rho').$$

经典曲棍球棒散度定义如下：

$$D_\alpha(P \parallel P') = \sum_i [P(i) - \alpha P'(i)]_+,$$

其中，$[x]_+ = \max(x,0)$。
我们从经典散度的定义开始证明：

$$\begin{aligned} D_\alpha(P \parallel P') &= \sum_i \max\left(0, \mathrm{Tr}(E_i\rho) - \alpha\mathrm{Tr}(E_i\rho')\right) \\ &= \sum_i \max\left(0, \mathrm{Tr}\left(E_i(\rho - \alpha\rho')\right)\right) \\ &= \sum_i \max\left(0, \mathrm{Tr}(E_i A)\right) \end{aligned}$$

对于任意半正定算子 $E_i$ 以及任意厄米算子 $A$，有 $\mathrm{Tr}(E_i A) \le \mathrm{Tr}(E_i A_+)$。由于 $A_+$ 是半正定算子，$\mathrm{Tr}(E_i A_+)$为非负。因此，我们可以得出 $\max(0,\mathrm{Tr}(E_i A)) \le \mathrm{Tr}(E_i A_+)$。

将此不等式应用于我们的表达式，可得：

$$\begin{aligned} D_\alpha(P \parallel P') &\le \sum_i \mathrm{Tr}(E_i A_+) \\ &= \mathrm{Tr}\left(\sum_i E_i A_+\right) \\ &= \mathrm{Tr}\left(\left(\sum_i E_i\right) A_+\right) \\ &= \mathrm{Tr}(I \cdot A_+) \\ &= D_\alpha^{(q)}(\rho \parallel \rho'). \end{aligned}$$

□

引理 B.3 建立了在固定测量下经典与量子 hockey-stick 散度之间的依赖关系。因此，我们可以从后处理流程中消除显式的测量映射 $f_{\text{mea}}$。具体而言，我们有：

$$\mathrm{D}_\alpha\big(Q^{(\eta)}(y) \,\|\, Q^{(\eta)}(y')\big) \le \mathrm{D}_\alpha^{(q)}\big(f_{\text{dep}}^{(\eta)} \circ f_{\text{enc}}(y) \,\|\, f_{\text{dep}}^{(\eta)} \circ f_{\text{enc}}(y')\big) \tag{4}$$

**引理 B.4.** 给定一个去极化信道 $f_{dep}^{(\eta)}(\rho) = \eta \frac{I}{d} + (1-\eta)\rho$，对于 $\eta \in [0,1]$ 和 $\alpha \ge 1$，我们有：

$$D_\alpha^{(q)}(f_{dep}^{(\eta)}(\rho) \parallel f_{dep}^{(\eta)}(\rho')) \le \max\left\{0, (1-\alpha)\frac{\eta}{d} + (1-\eta) D_\alpha^{(q)}(\rho \parallel \rho')\right\}$$

证明。 定义算子：

$$U = f_{dep}^{(\eta)}(\rho) - \alpha f_{dep}^{(\eta)}(\rho) = (1-\eta)(\rho - \alpha\rho') + \eta(1-\alpha)\frac{I}{d}.$$

于是：

$$\mathrm{D}_\alpha^{(q)}(f_{dep}^{(\eta)}(\rho) \parallel f_{dep}^{(\eta)}(\rho')) = \mathrm{Tr}[U_+],$$

其中， $U_+$ 表示 $U$的正部。

令 $P_+$ 为投影算子，其作用于 的正特征子空间。由于 $\mathrm{D}_\alpha^{(q)}(f_{(\eta)\ dep}(\rho) \parallel f_{dep}^{(\eta)}(\rho') > 0$，我们有 $\mathrm{Tr}[P_+] \ge 1$。 n:

$$\begin{aligned} \mathrm{Tr}[U_+] &= \mathrm{Tr}[P_+ U] \\ &= (1-\eta)\mathrm{Tr}[P_+(\rho - \alpha\rho')] + (1-\alpha)\frac{\eta}{d}\mathrm{Tr}[P_+] \\ &\le (1-\eta)\mathrm{D}_\alpha^{(q)}(\rho\|\rho') + (1-\alpha)\frac{\eta}{d}, \end{aligned}$$

由于 $\mathrm{Tr}[P_+] \ge 1$ 且 $1-\alpha \le 0$。 □

引理 B.4证明，量子曲棍球棒散度经过退极化信道后会收缩，收缩因子为（$1-\eta$），并带有一个取决于 $\alpha$ 和维度 $d$的加性项。将此结果应用于 $\rho = f_{\text{enc}}(y)$ 和 $\rho' = f_{\text{enc}}(y')$ 可得方程 4右端的一个上界，进而为方程 3给出一个界。

**定理 4.1** （失败概率放大）。 设 $A$： $\mathbb{X} \to \mathcal{P}(\mathbb{Y})$ 为一个满足 $(\varepsilon,\ \delta)$-差分隐私的经典机制，其中 $A = f_{par} \circ f_{cdp}$，并设 $Q^{(\eta)}$： $\mathbb{Y} \to \mathcal{P}(\mathbb{Z})$ 为 $d$ 维希尔伯特空间中的量子机制，定义为$Q^{(\eta)} = f_{mea} \circ f^{(\eta)}{}_{dep} \circ f_{enc}$，其中 $0 \le \eta \le 1$ 为退极化噪声因子。则复合机制 $Q^{(\eta)} \circ A$ 满足 $(\varepsilon',\ \delta')$-差分隐私，其中

$$\varepsilon' = \varepsilon, \quad \delta' = \left[\frac{\eta(1-e^{\varepsilon})}{d} + (1-\eta)\delta\right]_+$$

证明。设 $\mu = A(x)$ 与 $\nu = A(x')$ 为机制 $A$ 在相邻输入 $x$ 与 $x'$上的输出分布。我们旨在对曲棍球棒散度给出上界

$$\mathrm{D}_{e^\varepsilon}(\mu Q^{(\eta)} \| \nu Q^{(\eta)}).$$

根据引理 B.1，我们可以利用参数 $\mu$ 和 $\nu$ 将其分解，并定义辅助分布 $\theta = \mathrm{D}_{e^\varepsilon}(\mu\|\nu)$ $\mu'$、$\nu'$和$\omega$，使得 $\mu' \perp \nu'$

$$\mu = (1-\theta)\omega + \theta\mu', \quad \nu = \frac{1-\theta}{e^\varepsilon}\omega + \left(1 - \frac{1-\theta}{e^\varepsilon}\right)\nu'.$$

此外，定义 $\tilde{\varepsilon} = \log\left(1 + \frac{e^\varepsilon - 1}{\theta}\right)$。根据引理 B.1，可得

$$\mathrm{D}_{e^\varepsilon}(\mu\|\nu) \le \theta \cdot \mathrm{D}_{e^{\tilde{\varepsilon}}}(\mu'\|\nu').$$

现在考虑后处理输出：

$$\begin{aligned}
&\mathrm{D}_{e^\varepsilon}(\mu Q^{(\eta)} \| \nu Q^{(\eta)}) \\
&\le \theta \cdot \mathrm{D}_{e^{\tilde{\varepsilon}}}(\mu' Q^{(\eta)} \| \nu' Q^{(\eta)}) \\
&\le \theta \cdot \sup_{y \neq y'} \mathrm{D}_{e^{\tilde{\varepsilon}}}(Q^{(\eta)}(y) \| Q^{(\eta)}(y')) \text{(Lemma B.2)} \\
&\le \theta \cdot \sup_{y \neq y'} \mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}\left(f^{(\eta)}_{\text{dep}} \circ f_{\text{enc}}(y) \,\|\, f^{(\eta)}_{\text{dep}} \circ f_{\text{enc}}(y')\right) \text{(Lemma B.3)} \\
&= \theta \cdot \sup_{\rho, \rho'} \mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}\left(f^{(\eta)}_{\text{dep}}(\rho) \,\|\, f^{(\eta)}_{\text{dep}}(\rho')\right) \\
&\le \theta \cdot \max\left\{0, \frac{\eta(1-e^{\tilde{\varepsilon}})}{d} + (1-\eta) \cdot \mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}(\rho\|\rho')\right\} \text{(Lemma B.4)} \\
&\le \max\left\{0, \frac{\theta\eta(1-e^{\tilde{\varepsilon}})}{d} + \theta(1-\eta)\right\} \text{(Because } \mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}(\rho\|\rho') \le 1\text{)}
\end{aligned}$$

回顾一下，$e^{\tilde{\varepsilon}} = 1 + \frac{e^\varepsilon - 1}{\theta}$，我们将其代入表达式：

$$\begin{aligned}
\frac{\theta\eta(1-e^{\tilde{\varepsilon}})}{d} &= \frac{\theta\eta}{d}\left(1 - \left(1 + \frac{e^\varepsilon - 1}{\theta}\right)\right) \\
&= \frac{\theta\eta}{d}\left(-\frac{e^\varepsilon - 1}{\theta}\right) \\
&= \frac{\eta(1-e^\varepsilon)}{d}
\end{aligned}$$

此外，由于原始机制 $A$ 是 $(\varepsilon, \delta)$-差分隐私，因此有 $\theta = \mathrm{D}_{e^\varepsilon}(\mu\|\nu) \le \delta$。由于 $1 - \eta \ge 0$，得到最终结果：

$$\mathrm{D}_{e^\varepsilon}(\mu Q^{(\eta)} \| \nu Q^{(\eta)}) \le \left[\frac{\eta(1-e^\varepsilon)}{d} + (1-\eta)\delta\right]_+$$

□

定理 4.1 给出了 $\delta'$ 复合机制的失败概率 $Q^{(\eta)} \circ A$上界，同时将隐私损失固定为 $\varepsilon' = \varepsilon$。该结果是通过依次将引理 B.2 、 B.3 和 B.4 应用于公式 2 而推导出的。由最终上界可知，对于 $\varepsilon \in [0, 1]$，有 $\delta' \le \delta$。因此，失败概率严格降低，从而产生隐私放大效应，正如推论 4.2中所正式陈述的那样。

基于（Lecuyer 等人，2019年），我们在推论 $Q^{(\eta)} \circ A$ 中推导出了复合机制 $Q^{(\eta)} \circ A$ 的可认证对抗鲁棒性的显式条件 4.3。该条件定义了一个鲁棒性阈值，模型的期望置信度分数必须超过该阈值。值得注意的是，由于推论 4.2 中形式化的隐私放大效应，复合机制（由 $\delta'$ 参数化）下的鲁棒性阈值严格低于原始经典机制（由 $\delta$ 参数化）下的阈值。因此，量子后处理可证明扩大了能够保证对抗鲁棒性的输入集合。有关对抗鲁棒性的更多详情，请参阅附录 A。

**推论 4.2**。 复合机制 $Q^{(\eta)} \circ A$ 满足$(\varepsilon, \delta')$-差分隐私，且 $\delta' < \delta$，从而严格放大了整体失败概率。

证明。 目标是证明 $\delta' < \delta$ 对于量子后处理处于激活状态的任何非平凡情形（$\eta > 0$）。根据定理4.1，我们有：

$$\delta' = \left[\frac{\eta(1-e^{\varepsilon})}{d} + (1-\eta)\delta\right]_+$$

令第一项为 $C = {}^{\eta}{}_{(1-e^{\varepsilon})}\, d$。由于 $\eta > 0$、$d \geq 2$ 和 $\varepsilon > 0$，我们有 $C \leq 0$。由于 $C$ 严格为负，$C + (1-\eta)\delta < (1-\eta)\delta \leq \delta$。因此，$\delta' < \delta$。

□

**推论 4.3**。 若正确类别 $k$ 满足以下条件，则复合机制 $C = Q^{(\eta)} \circ A$ 对输入$x \in \mathbb{X}$ 具有可认证的对抗扰动鲁棒性：

$$\mathbb{E}[[C(x)]_k] > e^{2\varepsilon} \max_{i \neq k} \mathbb{E}[[C(x)]_i] + (1 + e^{\varepsilon})\delta'$$

证明。 相关研究见（Lecuyer 等人，2019年）。 □

### B.2. 定理 4.4 的证明及支撑引理

我们研究复合机制 $Q^{(\eta)} \circ A$ 如何同时放大隐私损失 $\varepsilon$ 和失败概率 $\delta$。我们的方法基于高级联合凸性理论，该理论最初由（Balle 等人，2018）提出。我们在下文将该理论重述为引理B.5。

**引理 B.5** （高级联合凸性）。设 $\mu, \mu'$ 为概率分布，满足

$$\mu = (1-\sigma)\mu_0 + \sigma\mu_1, \quad \mu' = (1-\sigma)\mu_0 + \sigma\mu_1',$$

对于某个 $\sigma \in [0,1]$ 以及分布 $\mu_0, \mu_1, \mu_1'$。给定 $\alpha \geq 1$，定义 $\alpha' = 1 + \sigma(\alpha - 1)$ $\beta = \frac{\alpha'}{\alpha}$。则以下不等式成立：

$$D_{\alpha'}(\mu \| \mu') \leq (1-\beta)\sigma D_{\alpha}(\mu_1 \| \mu_0) + \beta\sigma D_{\alpha}(\mu_1 \| \mu_1')$$

证明。见（Balle 等人，2018） □

引理 B.5 给出了散度 $D_{\alpha'}(\mu \| \mu')$ 的一个上界，该上界由分量分布 $\mu_0, \mu_1$ 之间的 $D_\alpha$ 个散度表示。随着混合参数 $\sigma$ 所控制的共享（重叠）分布 $\mu_0$ 的贡献增大，该上界会变得更紧。回到我们的分析，对于相邻输入 $x \sim x'$ 的 $\mu = A(x)$ 和 $\nu = A(x')$，以及更紧的隐私损失 $\varepsilon' = \log[1 + \sigma(e^{\varepsilon} - 1)]$，我们能够通过识别分布 $\mu Q^{(\eta)}$ 和 $\nu Q^{(\eta)}$ 之间的共享分量，将 $D_{e^{\varepsilon'}}(\mu Q^{(\eta)} \| \nu Q^{(\eta)})$ 界定为上界，如引理 B.6 所示。

**引理 B.6**。 设 $\rho$ 是一个定义在 $d$维希尔伯特空间上的密度矩阵，并令

$$\rho' = f_{dep}(\rho) = \eta\frac{I}{d} + (1-\eta)\rho$$

为其退极化版本，其中 $0 \leq \eta \leq 1$。设 $\{E_k\}_{k=1}^{K}$ 是一个满足 $\sum_k E_k = I$ 的正算子值测度。则测量概率满足：

$$\zeta'(k) = \frac{\eta}{d}\operatorname{Tr}(E_k) + (1-\eta)\zeta(k),$$

其中 $\zeta' = f_{mea}(\rho')$ 和 $\zeta = f_{mea}(\rho)$，其中 $\zeta', \zeta \in \mathcal{P}(\mathbb{Z})$。

证明。 根据迹算子的线性性,

$$\begin{aligned}\zeta'(k) &= \mathrm{Tr}(E_k\rho') \\ &= \mathrm{Tr}\left(E_k\left(\eta\frac{I}{d}+(1-\eta)\rho\right)\right) \\ &= \eta\,\mathrm{Tr}\left(E_k\frac{I}{d}\right)+(1-\eta)\,\mathrm{Tr}(E_k\rho) \\ &= \frac{\eta}{d}\,\mathrm{Tr}(E_k)+(1-\eta)\zeta_k.\end{aligned}$$

□

引理 B.6 阐明了量子系统中的退极化噪声 $\mathcal{H}$ 如何影响由此产生的经典输出分布 $\mathbb{Z}$。具体而言，它表明，退极化条件下的测量分布是原始（无噪声）分布与最大混合态对应分布的凸组合，其中噪声强度 $\eta$ 决定了混合比率。

根据引理 B.6，我们可以相应地分解输出分布 $\mu Q^{(\eta)}$ 和 $\nu Q^{(\eta)}$。根据定义，量子机制 $Q^{(\eta)}$ 可以表示为两个机制的凸组合： $Q^{(0)}$ （不施加噪声）和 $Q^{(1)}$ （施加完全退极化噪声）。机制 $Q^{(1)}$ 是常值机制，因为它始终输出最大混合态的测量分布。也就是说，对于所有 $y\in\mathbb{Y}$，都有 $Q^{(1)}(y)(k)=\mathrm{Tr}(E_k)/d$，其中 $E_k$ 是第 $k$ 个 POVM 元素，且 $d=\dim(\mathcal{H})$。我们将此常值输出分布记为 $\zeta_{\mathrm{mix}}$。另一方面， $Q^{(0)}(y)$对应于无噪声分布 $\zeta$，而 $Q^{(\eta)}(y)$对应于引理 B.6 中定义的分布 $\zeta'$。利用该引理给出的分解，我们得到

$$Q^{(\eta)}(y)=\eta Q^{(1)}(y)+(1-\eta)Q^{(0)}(y),\quad \forall y\in\mathbb{Y}$$

利用 $Q^{(\eta)}$ 的线性性以及表示式$\mu=\sum_{y\in\mathrm{supp}(\mu)}\mu(y)\tau_y$ 和$\nu=\sum_{y\in\mathrm{supp}(\nu)}\nu(y)\tau_y$，我们得到$\mu Q^{(\eta)}=\eta\zeta_{\mathrm{mix}}+(1-\eta)\mu Q^{(0)}$, 以及 $\nu Q^{(\eta)}=\eta\zeta_{\mathrm{mix}}+(1-\eta)\nu Q^{(0)}$。

应用高级联合凸性理论（引理B.5）于 $\mu Q^{(\eta)}$ 和 $\nu Q^{(\eta)}$，其中 $\varepsilon'=\log\big(1+(1-\eta)(e^{\varepsilon}-1)\big)$以及 $\beta=e^{\varepsilon'-\varepsilon}$，我们有:

$$\mathrm{D}_{e^{\varepsilon'}}\left(\mu Q^{(\eta)}||\nu Q^{(\eta)}\right)\leq(1-\eta)\left((1-\beta)\mathrm{D}_{e^{\varepsilon}}(\mu Q^{(0)}||\zeta_{\mathrm{mix}})+\beta\mathrm{D}_{e^{\varepsilon}}(\mu Q^{(0)}||\nu Q^{(0)})\right) \tag{5}$$

**引理 B.7**。给定最大混合态的测量分布 $\zeta_{mix}$ 以及任意分布 $z\in\mathcal{P}(\mathbb{Z})$，我们有:

$$D_\alpha(z||\zeta_{mix})\leq 1-\alpha\min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$$

证明。回忆曲棍球棒散度的定义:

$$\mathrm{D}_\alpha(z\|\zeta_{\mathrm{mix}})=\sum_k[z(k)-\alpha\zeta_{\mathrm{mix}}(k)]_+,$$

其中 $[x]_+=\max\{x,0\}$。由于 $\zeta_{\mathrm{mix}}(k)=\frac{\mathrm{Tr}(E_k)}{d}\geq\varphi=\min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$，我们有

$$[z(k)-\alpha\zeta_{\mathrm{mix}}(k)]_+\leq[z(k)-\alpha\varphi]_+.$$

对 $k$ 求和可得

$$\mathrm{D}_\alpha(z\|\zeta_{\mathrm{mix}})\leq\sum_k[z(k)-\alpha\varphi]_+.$$

由于 $\sum_k z(k)=1$，因此可得

$$\sum_k[z(k)-\alpha\varphi]_+\leq 1-\alpha\varphi.$$

因此，

$$\mathrm{D}_\alpha(z\|\zeta_{\text{mix}}) \leq 1-\alpha \min_k \left(\frac{\mathrm{Tr}(E_k)}{d}\right).$$

□

根据引理B.7，我们可以根据 POVM 元素的迹值，推导出 $\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}\|\zeta_{\text{mix}})$ 的一个上界。此外，根据曲棍球棒散度的数据处理不等式，我们有 $\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}\|\nu Q^{(0)}) \leq \mathrm{D}_{e^\varepsilon}(\mu\|\nu) \leq \delta$。综合这些结果，我们得到方程5的一个改进上界：

$$\mathrm{D}_{e^{\varepsilon'}}\left(\mu Q^{(\eta)}\|\nu Q^{(\eta)}\right) \leq (1-\eta)\left(1-e^{\varepsilon'-\varepsilon}(1-\delta)-(e^\varepsilon-e^{\varepsilon'})\varphi\right)$$

，其中 $\varphi = \text{最小值}_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$。这一结果在定理4.4中得到形式化，该定理刻画了退极化噪声如何增强复合机制 $Q^{(\eta)} \circ A$的隐私保证。具体而言，该机制满足$(\varepsilon',\ \delta')$-差分隐私，其中$\varepsilon' = \log\left(1+(1-\eta)(e^\varepsilon-1)\right)$ 且 $\delta' = (1-\eta)\left[1-e^{\varepsilon'-\varepsilon}(1-\delta)-(e^\varepsilon-e^{\varepsilon'})\varphi\right]$.

定理 4.4 表明，放大后的失败概率 $\delta'$ 取决于正算子值测度的选择。具体而言，$\delta'$ 会随着 $\varphi = \min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$ 的增大而收紧。这一发现引出了推论 4.5，并指出当所有 POVM 元素 $\delta'$ 的迹相等时，$E_k$ 达到最小值（即 $\mathrm{Tr}(E_k) = \frac{1}{K}$）。

相反，$\varepsilon' \leq \varepsilon$ 对于所有 $\eta \in [0,1]$，以 $\varepsilon$ 衡量的隐私损失始终会降低。然而，只有当噪声水平 $\delta$ 超过推论 $\delta' \leq \delta$给出的阈值时，$\eta$ 的界才会改善（即 4.6）。这一条件表明，只有达到足够的量子噪声水平，才能在两个参数上都严格放大隐私保证。

**定理 4.4**（隐私损失放大）。设 $A = f_{par} \circ f_{cdp}$ 是 $(\varepsilon, \delta)$-差分隐私，且 $Q^{(\eta)} = f_{mea} \circ f^{(\eta)}_{dep} \circ f_{enc}$ 是一个作用于 $d$ 维希尔伯特空间中的量子机制，其中 $0 \leq \eta \leq 1$ 为退极化噪声因子。那么，复合 $Q^{(\eta)} \circ A$ 是 $(\varepsilon', \delta')$- 差分隐私，其中 $\varepsilon' = \log\left(1+(1-\eta)(e^\varepsilon-1)\right)$，且 $\delta' = (1-\eta)\left(1-e^{\varepsilon'-\varepsilon}(1-\delta)-(e^\varepsilon-e^{\varepsilon'})\varphi\right)$，其中 $\varphi = \min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right)$。

证明。设 $x$、$x' \in \mathbb{X}$ 是相邻输入，即 $x \simeq x'$。设 $\mu = A(x)$ 和 $\nu = A(x')$ 分别表示 $A$ 的输出分布。根据定义，我们有 $Q^{(0)}$ 和 $Q^{(1)}$，它们分别是无噪声机制和完全噪声机制。可以看出，$Q^{(1)}$ 是一个常值机制，因为 $Q^{(1)}$ 的输出始终是对最大混合态的测量，即 $Q^{(1)}(y)(k) = \mathrm{Tr}(E_k)/d$，其中 $\forall y \in \mathbb{Y}$。根据引理 B.6，我们有：

$$Q^{(\eta)}(y) = \eta Q^{(1)}(y) + (1-\eta)Q^{(0)}(y), \forall y \in \mathbb{Y}$$

。因此，我们可以将 $Q^{(\eta)}$ 写成 $Q^{(0)}$ 和 $Q^{(1)}$ 的混合，其中 $Q^{(\eta)} = \eta Q^{(1)} + (1-\eta)Q^{(0)}$ .

令常值输出为 $Q^{(1)}$，即 $\zeta_{\text{mix}}$。根据（Balle 等人，2018）中的高级联合凸性定理，给定$\varepsilon' = \text{对数}\left(1+(1-\eta)(e^\varepsilon-1)\right)$，我们有：

$$\begin{aligned}
&\mathrm{D}_{e^{\varepsilon'}}\left(\mu Q^{(\eta)}||\nu Q^{(\eta)}\right)\\
&= \mathrm{D}_{e^{\varepsilon'}}\left(\eta\mu Q^{(1)} + (1-\eta)\mu Q^{(0)}||\eta\nu Q^{(1)} + (1-\eta)\nu Q^{(0)}\right)\\
&= \mathrm{D}_{e^{\varepsilon'}}\left(\eta\zeta_{\text{mix}} + (1-\eta)\mu Q^{(0)}||\eta\zeta_{\text{mix}} + (1-\eta)\nu Q^{(0)}\right)\\
&= (1-\eta)\mathrm{D}_{e^\varepsilon}\left(\mu Q^{(0)}||(1-\beta)\zeta_{\text{mix}} + \beta\nu Q^{(0)}\right)\\
&\text{(Based on the advanced joint convexity theorem, } \beta = e^{\varepsilon'-\varepsilon})\\
&\leq (1-\eta)\left((1-\beta)\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}||\zeta_{\text{mix}}) + \beta\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}||\nu Q^{(0)})\right)
\end{aligned}$$

我们有 $\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}||\zeta_{\text{mix}}) \leq 1-e^\varepsilon \min_k\left(\frac{\mathrm{Tr}(E_k)}{d}\right) = 1-e^\varepsilon\varphi$ 且 $\mathrm{D}_{e^\varepsilon}(\mu Q^{(0)}||\nu Q^{(0)}) \leq \mathrm{D}_{e^\varepsilon}(\mu||\nu) \leq \delta$。因此，我们可以得出结论：

$$\mathrm{D}_{e^{\varepsilon'}}\left(\mu Q^{(\eta)}||\nu Q^{(\eta)}\right) \leq (1-\eta)\left(1-e^{\varepsilon'-\varepsilon}(1-\delta)-(e^\varepsilon-e^{\varepsilon'})\varphi\right)$$

□

**推论 4.5**。 设 $\{E_k\}_{k=1}^K$ 为 $f_{mea}$中使用的正算子值测度。那么，定理 $\delta'$ 4.4 中的放大后的失败概率在所有 *POVM* 元素的迹相等时达到最小（即，$\mathrm{Tr}[E_k] = \frac{d}{K}$对所有 $k \in \{1, \ldots, K\}$均成立）。

证明。 目标是最小化放大后的失败概率 $\delta'$ 关于正算子值测度的选择 $\{E_k\}_{k=1}^K$。根据定理4.4， $\delta'$ 的表达式为:

$$\delta' = (1-\eta)\big(1 - e^{\varepsilon'-\varepsilon}(1-\delta) - (e^{\varepsilon} - e^{\varepsilon'})\varphi\big)$$

此表达式中的所有项都与具体的测量选择无关，只有 $\varphi = \min_k \left( \frac{\mathrm{Tr}(E_k)}{d} \right)$ 除外。

为了分析 $\delta'$ 如何依赖于 $\varphi$，我们考察 $-(1-\eta)(e^{\varepsilon} - e^{\varepsilon'})$ 的符号。由于 $\eta \in [0,1]$ 且 $\varepsilon' \le \varepsilon$，该系数非正。因此， $\delta'$ 是关于 $\varphi$ 的单调递减函数。

因此，要最小化 $\delta'$，我们必须最大化 $\varphi$。这等价于在满足正算子值测度完备性约束 $\sum_{k=1}^K E_k = I$ 的条件下，最大化 $\min_k(\mathrm{Tr}(E_k))$。对完备性关系两边取迹可得:

$$\sum_{k=1}^{K} \mathrm{Tr}(E_k) = \mathrm{Tr}(I) = d$$

当所有 $\mathrm{Tr}[E_k]$ 都相等时，函数 $\min_k(\mathrm{Tr}[E_k])$ 取得最大值。因此，最优选择是对所有 $k$均取 $\mathrm{Tr}[E_k] = d/K$。

□

**推论 4.6.** 给定一个满足 $\mathrm{Tr}[E_k] = \frac{d}{K} \forall k$ 的最优测量，复合机制 $Q^{(\eta)} \circ A$ 会严格改善隐私保证（即 $\varepsilon' \le \varepsilon$和 $\delta' \le \delta$），如果

$$\eta \ge 1 - \frac{\delta}{(1-\delta)(1-e^{-\varepsilon}) - (e^{\varepsilon}-1)/K}$$

证明. 我们求出在最优测量假设下确保 $\delta' \le \delta$ 的 $\eta$ 条件。根据推论 4.5，我们有 $\varphi = 1/K$。保证 $\varepsilon' \le \varepsilon$ 对所有 $\eta \in [0,1]$ 均成立。

我们从以下不等式开始 $\delta' \le \delta$，其中使用了定理4.4中的表达式:

$$(1-\eta)\big(1 - e^{\varepsilon'-\varepsilon}(1-\delta) - (e^{\varepsilon} - e^{\varepsilon'})\varphi\big) \le \delta$$

代入恒等式 $e^{\varepsilon'-\varepsilon} = 1 - \eta + \eta e^{-\varepsilon}$, $e^{\varepsilon} - e^{\varepsilon'} = \eta(e^{\varepsilon} - 1)$ 和 $\varphi = 1/K$，我们得到 e:

$$(1-\eta)\left(1 - (1-\eta+\eta e^{-\varepsilon})(1-\delta) - \frac{\eta(e^{\varepsilon}-1)}{K}\right) \le \delta$$

主括号内的表达式可简化为 $\delta + \eta(1-\delta)(1-e^{-\varepsilon}) - {}^{\eta}\!/_{(e^{\varepsilon}-1)\,K}$。将其代回并展开、整理 $\eta > 0$ 后，我们得到:

$$(1-\delta)(1-e^{-\varepsilon}) - \frac{e^{\varepsilon}-1}{K} - \delta \le \eta\left((1-\delta)(1-e^{-\varepsilon}) - \frac{e^{\varepsilon}-1}{K}\right)$$

解出 $\eta$ 即可得到阈值:

$$\begin{aligned}\eta &\ge \frac{(1-\delta)(1-e^{-\varepsilon}) - (e^{\varepsilon}-1)/K - \delta}{(1-\delta)(1-e^{-\varepsilon}) - (e^{\varepsilon}-1)/K} \\ &= 1 - \frac{\delta}{(1-\delta)(1-e^{-\varepsilon}) - (e^{\varepsilon}-1)/K}\end{aligned}$$

□

### B.3. 定理 4.10 的证明及支撑引理

在此，我们建立了一个严谨的框架，用于研究效用损失，其定义为机制的带噪版本与无噪版本之间的绝对误差。由于采用基于采样的测量过程，机制的最终输出具有随机性。因此，我们分析两者输出的期望值之差。期望值体现了机制的平均行为，并提供了一个确定性量，可用于衡量效用损失。

形式化地，我们将期望测量函数定义为 $f_{\text{exp}}$： $\mathcal{H} \to \mathbb{R}$如下：

$$f_{\text{exp}}(\rho) = \sum_k k \operatorname{Tr}[E_k \rho] = \operatorname{Tr}\left[\left(\sum_k k E_k\right)\rho\right] = \operatorname{Tr}[E_{\text{exp}}\rho]$$

其中， $E_{\text{exp}} = \sum_k k E_k$ 是期望值可观测量。

使用此函数，我们定义确定性期望机制。**完整机制**（包括经典噪声和量子噪声）为 $\mathcal{M}_{\text{full}}(x) = (f_{\text{exp}} \circ f^{(\eta)}$ 依赖 $\circ f_{\text{enc}} \circ f_{\text{par}} \circ f_{\text{cdp}})(x)$。另一方面，**无噪声机制（干净）**为 $\mathcal{M}_{\text{clean}}(x) = (f_{\text{exp}} \circ f_{\text{enc}} \circ f_{\text{par}})(x)$。总效用损失是它们期望输出之间的最坏情况绝对误差：

$$\text{Error} = \sup_{x \in \mathbb{X}} |\mathcal{M}_{\text{full}}(x) - \mathcal{M}_{\text{clean}}(x)|$$

为分析这一误差，我们引入一种**中间机制（半）**，它仅包含量子噪声，定义为 $\mathcal{M}_{\text{half}}(x) = (f_{\text{exp}} \circ f^{(\eta)}$ 依赖 $\circ f_{\text{enc}} \circ f_{\text{par}})(x)$。

**引理 B.8.** 中间机制 $\mathcal{M}_{half}$ 关于输入扰动 $\kappa$是 $L_\infty$-利普希茨的，满足$|\mathcal{M}_{half}(x+\kappa) - \mathcal{M}_{half}(x)| \le L_\infty \|\kappa\|_\infty$。 $L_\infty$ 定义如下：

$$L_\infty = 2(1-\eta)\|E_{exp}\|_{op}\|W\|_\infty \left(\sum_j \|H_j\|_{op}\right)$$

证明。 首先，我们证明，函数复合的利普希茨界可以由各函数利普希茨常数的乘积得到。具体而言，假设 $f$ 可以写成

$$f = f_1 \circ f_2 \circ \cdots \circ f_h,$$

其中， $\circ$表示函数复合，且每个 $f_i$ 都存在一个利普希茨常数 $L_i$，对应于 $i = 1, \ldots, h$。那么，对于任意输入 $x$和一个较小的偏差 $\kappa$，有

$$\begin{aligned}
&\|f(x+\kappa) - f(x)\| \\
&\le L_1 \|f_2 \circ \cdots \circ f_h(x+\kappa) - f_2 \circ \cdots \circ f_h(x)\| \\
&\le L_1 L_2 \|f_3 \circ \cdots \circ f_h(x+\kappa) - f_3 \circ \cdots \circ f_h(x)\| \\
&\vdots \\
&\le \left(\prod_{i=1}^{h} L_i\right)\|\kappa\|.
\end{aligned}$$

由于机制 $\mathcal{M}_{\text{half}}$ 表示为 $f_{\text{exp}}$、 $f^{\eta}$ dep、 $f_{\text{enc}}$和 $f_{\text{par}}$的复合，我们的目标是确定每个单独函数的利普希茨界。

**$f_{\text{par}}$的利普希茨界**：

函数 $f_{\text{par}}$： $\mathbb{X} \to \mathbb{Y}$ 定义为

$$f_{\text{par}}(x) = Wx + b,$$

‘由于 $b$ 是一个常数偏移（不会影响利普希茨连续性），我们有：

$$\|f_{\text{par}}(x+\kappa) - f_{\text{par}}(x)\| = \|W\kappa\| \leq \|W\|_\infty \|\kappa\|_\infty,$$

因此， $f_{\text{par}}$ 是 $\|W\|_\infty$-利普希茨的。

**$f_{\text{enc}}$的利普希茨界**：

函数 $f_{\text{enc}}$： $\mathbb{Y} \to \mathcal{H}$ 将经典向量 $y$ 编码为密度矩阵 $f_{\text{enc}}(y) = U_{\text{enc}}(y)|0\rangle\langle 0|U_{\text{enc}}(y)^\dagger$， 其中$U_{\text{enc}}(y) = \prod_{j=1}^{N} e^{-i(\mathbf{w}_j \cdot y_j + b_j)H_j}$。我们需要用 $\|\kappa\|_\infty$来界定迹范数距离 $f_{\text{enc}}(y+\kappa) - f_{\text{enc}}(y)$。

$$\begin{aligned}
&\|f_{\text{enc}}(y+\kappa) - f_{\text{enc}}(y)\| \\
&= \|U_{\text{enc}}(y+\kappa)\rho_0 U_{\text{enc}}(y+\kappa)^\dagger - U_{\text{enc}}(y)\rho_0 U_{\text{enc}}(y)^\dagger\| \\
&\leq 2\|U_{\text{enc}}(y+\kappa) - U_{\text{enc}}(y)\|
\end{aligned}$$

其中 $\rho_0 = |0\rangle\langle 0|$， 并且我们使用了三角不等式和迹范数的性质。酉算子之差的界为：

$$\begin{aligned}
&\|U_{\text{enc}}(y+\kappa) - U_{\text{enc}}(y)\| \\
&\leq \sum_{j=1}^{N} \|e^{-i(y_j+\kappa_j)H_j} - e^{-iy_jH_j}\| \\
&\leq \sum_{j=1}^{N} |\kappa_j| \|H_j\| \leq \sum_{j=1}^{N} \|H_j\| \|\kappa\|_\infty
\end{aligned}$$

(Based on (Berberich et al., 2024))

因此， $f_{\text{enc}}$ 是 $2\left(\sum_{j=1}^{n} \|H_j\|\right)$-利普希茨的。

$f^{(\eta)}$ **退极化** 的利普希茨界：

函数 $f^{(\eta)}_{\text{退极化}}$： $\mathcal{H} \to \mathcal{H}$ 表示退极化噪声：

$$f^{(\eta)}_{\text{dep}}(\rho) = (1-\eta)\rho + \eta\frac{I}{d},$$

其中， $I$ 是单位矩阵， $d$是希尔伯特空间的维数。

由于项 $\eta\frac{I}{d}$ 是常数，两个输出之间的差为：

$$\|f^{(\eta)}_{\text{dep}}(\rho) - f^{(\eta)}_{\text{dep}}(\sigma)\| = (1-\eta)\|\rho - \sigma\|.$$

因此， $f^{(\eta)}_{\text{退极化}}$ 是 $(1-\eta)$-利普希茨函数。

**利普希茨界 $f_{\text{exp}}$**：

测量函数 $f_{\text{mea}}$： $\mathcal{H} \to \mathbb{R}^K$ 由一组正算子值测度 $\{E_k\}$ 定义，将量子态 $\rho$ 映射为概率向量：

$$f_{\text{exp}}(\rho) = \sum_k k \operatorname{Tr}(E_k\rho).$$

给定 $E_{\text{exp}} = \sum_k kE_k$，根据迹对偶性和 H¨older 不等式，我们有：

$$\|f_{\text{exp}}(\rho) - f_{\text{exp}}(\rho')\| = \left|\operatorname{Tr}(E_{\text{exp}}(\rho - \rho'))\right| \leq \|E_{\text{exp}}\|_{op}\|\rho - \rho'\|.$$

因此， $f_{\text{exp}}$ 是 $\|E_{\text{exp}}\|_{op}$-利普希茨的。

因此，机制 $\mathcal{M}_{\text{half}}$ 是 $L_\infty$-利普希茨的，其中 $L_\infty = 2(1-\eta)\|E_{\text{exp}}\|_{op}\|W\|_\infty \left(\sum_j \|H_j\|_{op}\right)$ .

□

引理 B.8 给出了 $\mathcal{M}_{\text{half}}$ 关于其经典输入扰动的敏感度上界。我们用$\|\cdot\|_p$ 表示 $p$-范数，用 $\|\cdot\|_{op}$ 表示算子范数。证明利用利普希茨连续性的链式法则，其中总体利普希茨常数 $L_\infty$ 等于复合中各分量函数对应常数的乘积，即 $f_{\text{exp}}$， $f_{\text{dep}}$， $f_{\text{enc}}$，以及 $f_{\text{par}}$。此外，我们注意到，若 $\kappa \sim \mathcal{N}(0,\ \sigma^2 I)$，则 $\mathcal{M}_{\text{half}}(x+\kappa)$ 与 $\mathcal{M}_{\text{full}}(x)$的分布相同。因此，该引理给出了这两个机制之间差异的上界。

**引理 B.9.** 中间机制与无噪声机制的期望输出之差的绝对值一致有界如下：

$$|\mathcal{M}_{half}(x) - \mathcal{M}_{clean}(x)| \leq 2\eta\|E_{exp}\|_{op}$$

证明。 设 $\rho(x) = (f_{\text{enc}} \circ f_{\text{par}})(x)$ 为干净的量子态。

$$\begin{aligned}
&|\mathcal{M}_{\text{half}}(x) - \mathcal{M}_{\text{clean}}(x)| \\
&= \left|\text{Tr}[f_{\text{exp}} \cdot f^{\eta}_{\text{dep}}(\rho(x))) - \text{Tr}(f_{\text{exp}} \cdot \rho(x))\right| \\
&\leq \|E_{\text{exp}}\|_{op} \cdot \|f^{\eta}_{\text{dep}}(\rho(x)) - \rho(x)\| \quad \text{(Lipschitz property of } f_{\text{exp}})
\end{aligned}$$

迹距离项的界为：

$$\|f^{\eta}_{\text{dep}}(\rho) - \rho\| = \|((1-\eta)\rho + \eta\frac{I}{d}) - \rho\| = \eta\left\|\frac{I}{d} - \rho\right\|$$

由于 $\rho$ 和 $I/d$ 都是有效的密度矩阵，因此它们之间的迹距离至多为 2。因此， $\|\frac{I}{d} - \rho\| \leq 2$。将其代回即可得到最终界 $2\eta\|E_{\text{exp}}\|_{op}$。 □

引理 B.9 直接界定了 $\mathcal{M}_{\text{half}}$ 与 $\mathcal{M}_{\text{clean}}$之间的差异。该证明利用了函数 $f_{\text{exp}}$ 的利普希茨性质，以及任意两个密度矩阵之间的迹范数差至多为 2这一基本性质。结合引理 B.8中的结果，我们可以建立绝对误差的界。

**定理 4.10** （效用界）。 设作用于维数为 $d_X = \dim(\mathbb{X})$的输入空间 $\mathbb{X}$ 上的经典噪声为 $\kappa \sim \mathcal{N}$（$0$， $\sigma^2 I$）。对于任意给定的失败概率 $p > 0$，效用损失以概率方式界定为：

$$\Pr\left(\textit{Error} \leq L_\infty \cdot \sigma\sqrt{2\ln\frac{2d_X}{p}} + 2\eta\|E_{exp}\|_{op}\right) \geq 1 - p$$

其中 $L_\infty = 2\ (1-\eta)$
$\|E_{exp}\|_{op}\|W\|_\infty \left(\sum_j \|H_j\|_{op}\right)$。

证明。 对于给定的 $x$ 和经典噪声 $\kappa$，我们对绝对误差应用三角不等式 ：

$$\begin{aligned}
&|\mathcal{M}_{\text{full}}(x) - \mathcal{M}_{\text{clean}}(x)| \\
&= |\mathcal{M}_{\text{half}}(x+\kappa) - \mathcal{M}_{\text{clean}}(x)| \\
&\leq |\mathcal{M}_{\text{half}}(x+\kappa) - \mathcal{M}_{\text{half}}(x)| + |\mathcal{M}_{\text{half}}(x) - \mathcal{M}_{\text{clean}}(x)|
\end{aligned}$$

应用我们的两个 le mmas，第一项有界于 $L_\infty \cdot \|\kappa\|_\infty$ ，第二项有界于 $2\eta\|E_{\text{exp}}\|_{op}$ .

$$\text{Error} \leq L_\infty \cdot \|\kappa\|_\infty + 2\eta\|E_{\text{exp}}\|_{op}$$

随机误差取决于 $\|\kappa\|_\infty = \max_i |\kappa_i|$ 的大小，其中噪声向量的每个分量 $\kappa_i$ 都是从高斯分布中独立抽取的， $\kappa_i \sim \mathcal{N}(0, \sigma^2)$。

为了获得 $d$ 个独立高斯随机变量最大值的高概率界，我们可以对其尾部分布应用标准的并集界。对于任意给定的失败概率 $p > 0$，以至少 $1 - p$ 的概率， $\kappa$ 的无穷范数有如下界：

$$\|\kappa\|_\infty \leq \sigma\sqrt{2\ln(2d_X/p)}$$

结合这些界 ，我们可以得出，对于任意 $p > 0$，总效用损失以至少 $1 - p$ 的概率满足如下界 ：

$$\text{Error} \leq L_\infty \cdot \sigma\sqrt{2\ln\frac{2d_X}{p}} + 2\eta\|E_{\text{exp}}\|_{op}$$

□

定理 4.10结合了先前的结果，从而给出了单一的效用保证。证明利用三角不等式，将引理 B.8 和B.9中的界相加。由于经典噪声无界，最终的保证是一个高概率陈述，展示了效用损失与经典噪声（$\sigma$）和量子噪声（$\eta$）水平之间的权衡。

### B.4. 第三项分析——扩展至其他量子噪声信道（扩展版）

在本节中，除了给出定理4.7 和4.9的证明外，我们还进一步讨论如何将定理4.1 中的隐私放大结果推广至退极化噪声之外的广泛量子噪声信道类别。首先，我们找出实现隐私放大的关键机制。然后，通过分析两种非对称且具有实际物理意义的噪声过程，阐述这一推广：广义振幅阻尼（GAD）信道和广义退相位（GD）信道。

#### B.4.1. 推广背后的关键见解

这里，我们首先回顾附录B.1中给出的定理4.1的证明思路。分析首先利用引理B.1对机制在相邻输入上的输出分布进行分解。接着，利用引理B.2将散度分析归约为最坏情况下的一对正交输入。关键的是，引理B.3证明，测量结果的经典曲棍球棒散度不超过演化后的量子态的量子曲棍球棒散度。

我们可以看到，在定理4.1中，隐私放大是由引理B.4推导得出的，该引理证明了量子曲棍球棒散度 $D_\alpha^{(q)}$在去极化信道下的收缩。虽然定理4.1 利用了 $D_{e^\varepsilon}^{(q)}$ 的特定形式，但我们指出，即使将界放宽到标准迹距离 $D_1^{(q)}$ ，隐私保证仍然成立。˙具体而言，对于任意隐私参数 $\alpha \geq 1$ （其中 $\alpha = e^\varepsilon$ 在我们的语境下），量子曲棍球棒散度有如下上界，即迹距离散度：

$$D_\alpha^{(q)}(\rho\|\sigma) = \text{Tr}[(\rho - \alpha\sigma)_+] \leq \text{Tr}[(\rho - \sigma)_+] = D_1^{(q)}(\rho\|\sigma).$$

这个不等式成立，是因为减去更大的 $\sigma$ 倍数（因为 $\alpha \geq 1$）会缩小算子差的正部。

接下来，关键在于，要将定理4.1推广到任意噪声信道 $\mathcal{E}$，我们需要确定其在迹距离（或 $D_1$ 散度）下的收缩系数。如果信道 $\mathcal{E}$ 满足收缩界 $\kappa$（$\mathcal{E}$），使得：

$$\sup_{\rho\neq\sigma}\frac{D_1(\mathcal{E}(\rho)\|\mathcal{E}(\sigma))}{D_1(\rho\|\sigma)} \leq \kappa(\mathcal{E}),$$

那么，复合机制自然满足隐私放大，其中失败概率 $\delta$ 会按 $\kappa(\mathcal{E})$ 缩放。接下来的小节中，我们将这一见解应用于两种非对称噪声信道。

B.4.2. 广义振幅阻尼信道（定理 4.7 的证明）

广义振幅阻尼（GAD）信道是一种描述量子比特与其热环境之间能量交换的噪声过程。与退极化不同，GAD 本质上具有非对称性，因为它会驱使量子比特趋向一个依赖于温度的平衡态，同时抑制量子相干性。该信道由阻尼强度 $\eta \in [0,1]$ 和激发概率 $p \in [0,1]$参数化，其中 $p=0$ 对应趋向 $|0\rangle$的弛豫， $p=1$对应趋向 $|1\rangle$，中间值则代表非零温度行为。这种非对称性使 GAD 成为超导和囚禁离子器件的逼真噪声模型。为将其形式化，我们考虑作用于一个指定量子比特的 $n$量子比特信道：

$$f_{\mathrm{GAD}}^{(p,\eta)} = I_{2^{n-1}} \otimes A_{\mathrm{GAD}}^{(p,\eta)}.$$

尽管 GAD 信道具有非幺性质，但它会降低量子态之间的可区分性。由于所有量子态都会弛豫至同一个热固定点，激发概率的差异会缩小；与此同时，量子相干性的差异则会因能量耗散而衰减。在引理 B.10中，我们构造了 $f^{(p,\eta)}$ GAD的收缩系数。

**引理 B.10.** 设 $A^{(p,\eta)}$ 广义振幅阻尼为作用于单个量子比特的广义振幅阻尼（*GAD*）信道，其阻尼参数为$\eta \in [0,1]$，激发参数为 $p \in [0,1]$。定义 $n$量子比特信道

$$f_{\mathrm{GAD}}^{(p,\eta)} := \mathrm{I}_{2^{n-1}} \otimes A_{GAD}^{(p,\eta)},$$

则 $f^{(p,\eta)}$ 广义振幅阻尼的收缩系数满足

$$\kappa(f_{\mathrm{GAD}}^{(p,\eta)}) := \sup_{\rho \neq \sigma} \frac{D_1\left(f_{\mathrm{GAD}}^{(p,\eta)}(\rho) \,\middle\|\, f_{\mathrm{GAD}}^{(p,\eta)}(\sigma)\right)}{D_1(\rho\|\sigma)} \;\leq\; 2\sqrt{\eta} - \eta$$

证明。 根据定义 $D_1(\rho\|\sigma) = \frac{1}{2}\|\rho-\sigma\|_1$，因此 $\kappa(f^{(p,\eta)}$ 广义振幅阻尼$)$ 是信道$f^{(p,\eta)}$ 广义振幅阻尼 $= \mathrm{I}_{2^{n-1}} \otimes A^{(p,\eta)}$ 广义振幅阻尼的迹距离收缩系数：

$$\kappa\big(f_{\mathrm{GAD}}^{(p,\eta)}\big) = \sup_{\rho \neq \sigma} \frac{\|f_{\mathrm{GAD}}^{(p,\eta)}(\rho) - f_{\mathrm{GAD}}^{(p,\eta)}(\sigma)\|_1}{\|\rho-\sigma\|_1}.$$

上确界取遍所有 $n$量子比特态 $\rho, \sigma$，这些态可能在辅助系统和噪声量子比特之间存在纠缠。

根据（Hirche，2024年），该值上界为单量子比特信道 $A^{(p,\eta)}$ 广义振幅阻尼的完全迹距离收缩系数，其定义为

$$\eta_{Tr}^{c}(A_{\mathrm{GAD}}^{(p,\eta)}) := \sup_{k \geq 1} \sup_{\rho \neq \sigma} \frac{\|(\mathrm{I}_k \otimes A_{\mathrm{GAD}}^{(p,\eta)})(\rho) - (\mathrm{I}_k \otimes A_{\mathrm{GAD}}^{(p,\eta)})(\sigma)\|_1}{\|\rho-\sigma\|_1}.$$

由于对于某个特定的辅助系统维度， $f^{(p,\eta)}$ 广义振幅阻尼 恰好等于 $\mathrm{I}_{2^{n-1}} \otimes A^{(p,\eta)}$ 广义振幅阻尼，因此我们有

$$\kappa\big(f_{\mathrm{GAD}}^{(p,\eta)}\big) \leq \eta_{Tr}^{c}(A_{\mathrm{GAD}}^{(p,\eta)}).$$

根据（Hirche，2024年）中的引理 9 和命题 28，我们有：

$$\begin{aligned}
\kappa(f_{\mathrm{GAD}}^{(p,\eta)}) &= \eta_{Tr}^{c}(A_{\mathrm{GAD}}^{(p,\eta)}) \\
&\leq 1 - \alpha(A_{\mathrm{GAD}}^{(p,\eta)}) \\
&= 1 - (1-\sqrt{\eta})^2 \\
&= 1 - (1 - 2\sqrt{\eta} + \eta) \\
&= 2\sqrt{\eta} - \eta.
\end{aligned}$$

□

基于 L 引理 B.10，我们现在推导出定理 4. 7.

**定理 4.7** （广义振幅阻尼噪声下的放大）。 设 $A:\mathbb{X}\to\mathcal{P}(\mathbb{Y})$ 是满足 $(\varepsilon,\delta)$-差分隐私的经典机制，其中 $A=f_{par}\circ f_{cdp}$，并设 $Q^{(p,\eta)}:\mathbb{Y}\to\mathcal{P}(\mathbb{Z})$ 是在 $d$ 维希尔伯特空间中定义的量子机制，形式为 $Q^{(p,\eta)}=f_{mea}\circ f^{(p,\eta)}_{\text{广义振幅阻尼}}\circ f_{enc}$。那么，复合机制 $Q^{(p,\eta)}\circ A$ 满足 $(\varepsilon',\delta')$-差分隐私，其中

$$\varepsilon'=\varepsilon,\quad \delta'=(2\sqrt{\eta}-\eta)\delta.$$

证明。 设 $\mu=A(x)$ 和 $\nu=A(x')$ 分别为机制 $A$ 在相邻输入 $x$ 和 $x'$ 下的输出分布。我们旨在对曲棍球棒散度进行界定

$$\mathrm{D}_{e^\varepsilon}(\mu Q^{(p,\eta)}\|\nu Q^{(p,\eta)}).$$

根据引理 B.1，我们可以利用参数 $\mu$ 和 $\nu$ 对其进行分解，参数为 $\theta=\mathrm{D}_{e^\varepsilon}(\mu\|\nu)$，并定义辅助分布 $\mu',\nu'$ 和 $\omega$，满足 $\mu'\perp\nu'$，使得

$$\mu=(1-\theta)\omega+\theta\mu',\quad \nu=\frac{1-\theta}{e^\varepsilon}\omega+\left(1-\frac{1-\theta}{e^\varepsilon}\right)\nu'.$$

此外，定义 $\tilde{\varepsilon}=\log\!\left(1+\frac{e^\varepsilon-1}{\theta}\right)$。——

现在我们考虑后处理输出：

$$\begin{aligned}
&\mathrm{D}_{e^\varepsilon}(\mu Q^{(p,\eta)}\|\nu Q^{(p,\eta)})\\
&\le\theta\cdot\mathrm{D}_{e^{\tilde{\varepsilon}}}(\mu' Q^{(p,\eta)}\|\nu' Q^{(p,\eta)}) && \text{(引理 B.1)}\\
&\le\theta\cdot\sup_{y\neq y'}\mathrm{D}_{e^{\tilde{\varepsilon}}}(Q^{(p,\eta)}(y)\|Q^{(p,\eta)}(y')) && \text{(引理 B.2)}\\
&\le\theta\cdot\sup_{y\neq y'}\mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}\left(f^{(p,\eta)}_{\mathrm{GAD}}\circ f_{\mathrm{enc}}(y)\,\|\,f^{(p,\eta)}_{\mathrm{GAD}}\circ f_{\mathrm{enc}}(y')\right) && \text{(Lemma B.3)}\\
&=\theta\cdot\sup_{\rho,\rho'}\mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}\left(f^{(p,\eta)}_{\mathrm{GAD}}(\rho)\,\|\,f^{(p,\eta)}_{\mathrm{GAD}}(\rho')\right) && \text{(其中 } \rho,\ \rho' \text{ 均为纯态)}\\
&\le\theta\cdot\sup_{\rho,\rho'}\mathrm{D}^{(q)}_{1}\left(f^{(p,\eta)}_{\mathrm{GAD}}(\rho)\,\|\,f^{(p,\eta)}_{\mathrm{GAD}}(\rho')\right) && (e^{\tilde{\varepsilon}}\ge 1)\\
&\le\theta\cdot\sup_{\rho,\rho'}\left((2\sqrt{\eta}-\eta)\cdot\mathrm{D}^{(q)}_{1}(\rho\|\rho')\right) && \text{(引理 B.10)}\\
&\le\theta\cdot(2\sqrt{\eta}-\eta)\cdot 1 && (\text{因为 } \mathrm{D}^{(q)}_{e^{\tilde{\varepsilon}}}(\rho\|\rho')\le 1)
\end{aligned}$$

由于原始机制 $A$ 是 $(\varepsilon,\delta)$-差分隐私，因此我们有 $\theta=\mathrm{D}_{e^\varepsilon}(\mu\|\nu)\le\delta$。将其代入最终界：

$$\mathrm{D}_{e^\varepsilon}(\mu Q^{(p,\eta)}\|\nu Q^{(p,\eta)})\le(2\sqrt{\eta}-\eta)\cdot\delta.$$

由此得到所述的差分隐私参数。 □

### B.4.3. 广义退相位信道（定理 4.9 的证明）

广义退相位（GD）信道是量子信息中最基础且研究最广泛的噪声过程之一。它会抑制量子相干性，同时保持经典布居不变。具体而言，该信道可表述为：

$$A^{(\eta)}_{\mathrm{GD}}(\rho)=(1-\eta)\rho+\eta Z\rho Z$$

其中 $\eta\in[0,1]$ 是退相位参数，而 $Z$ 是泡利 Z 算符。根据（Hirche，2024年）中的命题 33，单量子比特 GD 通道的完全迹距离收缩系数恰好为 1（即 $\eta^c_{Tr}(A^{(\eta)}{}_{\mathrm{GD}})=1$）。这意味着在退相位噪声下，无法保证最坏情况下的隐私放大。换言之，对于某些输入态，噪声完全不会降低可区分性。不过，只有当状态之间的差异完全体现在对角差异上时，才会达到这种最坏情况。在许多常见的量子机器学习架构中，例如采用角度编码的架构，编码后的数据属于一类状态，其所有信息都包含在非对角分量（相干项）中。在这种情况下，GD 噪声确实会产生非平凡的收缩，因此能够实现隐私放大。假设4.8中正式介绍了这种情况的一个实例。

**假设 4.8**（乘积　　对所有量子比特进行赤道编码）。 对于每个输入 $y \in \mathbb{Y}$，编码器都会制备一个乘积态 $e$

$$\rho_y = f_{\text{enc}}(y) = \bigotimes_{j=1}^{n} \rho_y^{(j)},$$

其中每个单量子比特因子 $\rho_y^{(j)}$均为布洛赫球上的赤道态，即

$$\rho_y^{(j)} = \frac{1}{2}\Big(I + \cos\phi_y^{(j)}\, X + \sin\phi_y^{(j)}\, Y\Big),$$

对应于某个角度 $\phi_y^{(j)} \in \mathbb{R}$，且不含 $Z$分量。

这一赤道态假设可由常见的量子机器学习编码器满足，其中数据通过单量子比特旋转和哈达玛型制备被映射为相位与叠加，例如形式为 $H \to R_Z(\phi_y^{(j)})$ 的量子电路作用于每个量子比特（Schuld & Killoran，2019年；P´erez-Salinas 等人，2020年；Hatakeyama-Sato 等人，2023年）。除量子机器学习之外，赤道态在量子通信和量子密钥分发（QKD）中也起着核心作用，用于相位编码和基于相干性的信息传输（Fisher 等人，2014年；Xiao 等人，2014年）。因此，在该假设下分析 GD 通道的隐私放大既符合实际，也具有实践意义。

我们考虑作用于每个量子比特且彼此独立的 $n$量子比特 GD 通道，如下所示：

$$f_{\text{GD}}^{(\eta)} = \bigotimes_{j=1}^{n} A_{\text{GD}}^{(\eta)},$$

现在，我们将在假设 $f^{(\eta)}{}_{\text{GD}}$ 满足假设 4.8 的条件下，在引理 B.11中建立一个收缩界。

**引理 B.11**。 设 $f^{(\eta)}{}_{\text{GD}}$ 为上文定义的 $n$量子比特 *GD* 通道，并假设编码器 $f_{\text{enc}}$ 满足假设 *4.8*。则 $f^{(\eta)}{}_{\text{GD}}$ 在编码器族 $\{\rho_y\}_{y\in\mathbb{Y}}$ 上的迹距离收缩系数满足

$$\kappa\big(f_{\text{GD}}^{(\eta)}\big) := \sup_{y\neq y'} \frac{D_1\big(f_{\text{GD}}^{(\eta)}(\rho_y)\,\|\, f_{\text{GD}}^{(\eta)}(\rho_{y'})\big)}{D_1(\rho_y\|\rho_{y'})} \;\le\; |1-2\eta|.$$

证明。 对于每个 $y$，我们有

$$\rho_y = \bigotimes_{j=1}^{n} \rho_y^{(j)}, \qquad \rho_y^{(j)} = \tfrac{1}{2}\big(I + \cos\phi_y^{(j)} X + \sin\phi_y^{(j)} Y\big).$$

Let $\Delta = \rho_y - \rho_{y'}$ 对于两个不同的输入 $y \neq y'$。将 $\Delta$ 在 $n$量子比特泡利基下展开，可得：

$$\Delta = \sum_{P\in\mathcal{P}_n} c_P P,$$

其中，$\mathcal{P}_n = \{I, X, Y, Z\}^{\otimes n}$为 $n$量子比特泡利群，并且由于 $\Delta$是厄米的，有 $c_P \in \mathbb{R}$。

由于每个单量子比特因子 $\rho_y^{(j)}$仅包含 $I$、$X$ 和 $Y$ 分量，而不含 $Z$ 分量，因此任何乘积态$\rho_y = \bigotimes_j \rho_y^{(j)}$展开后都只包含单量子比特因子属于 $\{I, X, Y\}$的泡利串。对 $\rho_{y'}$同样如此，因此它们的差异 $\Delta = \rho_y - \rho_{y'}$ 在仅由 $I$和 $Z$组成的任何泡利串上都没有支撑。特别地，

$$c_P = 0 \quad \text{for all } P \in \mathcal{P}_n \text{ such that } P \in \{I, Z\}^{\otimes n}.$$

等价地，每个　　非零系数 $c_P$对应于一个泡利串 $P$，其中至少包含一个 $X$ 或 $Y$因子。

$n$ 量子比特 GD 通道在泡利基中对角作用：

$$f_{\text{GD}}^{(\eta)}(P) = \lambda_P P,$$

其中

$$\lambda_P = \prod_{j=1}^{n} \lambda_{P_j}, \quad \lambda_I = \lambda_Z = 1, \quad \lambda_X = \lambda_Y = 1 - 2\eta,$$

和 $P = P_1 \otimes \cdots \otimes P_n$ w ith $P_j \in \{I, X, Y, Z\}$。因此，对于任意至少包含一个 $X$ 或 $Y$ 的泡利串 $P$，我们有

$$\lambda_P = (1 - 2\eta)^k$$

对于 $k \geq 1$。因此，我们有：

$$|\lambda_P| \leq |1 - 2\eta|.$$

这意味着：

$$f_{\mathrm{GD}}^{(\eta)}(\Delta) = \sum_{P \in \mathcal{P}_n} c_P \lambda_P P,$$

其中每个非零系数均满足 $|\lambda_P| \leq |1 - 2\eta|$。作为定义在由至少包含一个 $X$ 或 $Y$ 的泡利串张成的子空间上的线性映射，$f^{(\eta)}$ 广义退相位 在正交归一算子基下为对角形式，其特征值的模不超过 $|1 - 2\eta|$。因此，它在该子空间上关于任意酉不变范数（尤其是迹范数）的算子范数至多为 $|1 - 2\eta|$。具体而言，

$$\|f_{\mathrm{GD}}^{(\eta)}(\Delta)\|_1 \leq |1 - 2\eta| \, \|\Delta\|_1.$$

由于 $D_1(\rho\|\sigma) = \frac{1}{2}\|\rho - \sigma\|_1$，我们得出，对于所有 $y \neq y'$，

$$\frac{D_1\big(f_{\mathrm{GD}}^{(\eta)}(\rho_y) \,\big\|\, f_{\mathrm{GD}}^{(\eta)}(\rho_{y'})\big)}{D_1(\rho_y \| \rho_{y'})} = \frac{\frac{1}{2}\|f_{\mathrm{GD}}^{(\eta)}(\Delta)\|_1}{\frac{1}{2}\|\Delta\|_1} \leq |1 - 2\eta|.$$

对所有 $y \neq y'$ 取上确界，即可得到所需的界。 □

最后，类似于退极化噪声和广义振幅阻尼噪声的隐私放大分析，我们现在推导适用于所有量子比特的 GD 通道的隐私放大定理。结合引理 B.11中的收缩界，以及前文使用的经典后处理和分布分解工具，即可立即得到该结果。

**定理 4.9.** 设 $A : \mathbb{X} \to \mathcal{P}(\mathbb{Y})$ 是一个满足 $(\varepsilon, \delta)$-差分隐私的经典机制，并设

$$Q^{(\eta)} := f_{\mathrm{mea}} \circ f_{\mathrm{GD}}^{(\eta)} \circ f_{\mathrm{enc}}$$

是一个 $n$-量子比特量子机制，其中 $f^{(\eta)}$ GD 是上述定义的 $n$-量子比特 ***GD*** 通道，并且 $f_{\mathrm{enc}}$ 满足假设 *4.8*。那么，复合机制 $Q^{(\eta)} \circ A$ 满足 $(\varepsilon', \delta')$-差分隐私，其中

$$\varepsilon' = \varepsilon, \qquad \delta' = |1 - 2\eta| \cdot \delta.$$

证明。 设 $\mu = A(x)$ 和 $\nu = A(x')$ 分别是 $A$ 在相邻输入 $x, x'$ 上的输出分布。如定理 4.7中一样，我们应用引理 B.1 ，利用参数 $\mu, \nu$ 对其进行分解，并将分析简化为最坏情况下的一对正交输入。利用引理B.2 和引理 B.3，我们可以界定

$$D_{e^\varepsilon}(\mu Q^{(\eta)} \| \nu Q^{(\eta)}) \leq \theta \cdot \sup_{\rho \neq \rho'} D_1\big(f_{\mathrm{GD}}^{(\eta)}(\rho) \,\big\|\, f_{\mathrm{GD}}^{(\eta)}(\rho')\big),$$

其中，上确界取遍 $f_{\mathrm{enc}}$ 的像中的编码态 $\rho$， $\rho'$。

根据引理 B.11，

$$D_1\big(f_{\mathrm{GD}}^{(\eta)}(\rho) \,\big\|\, f_{\mathrm{GD}}^{(\eta)}(\rho')\big) \leq |1 - 2\eta| \, D_1(\rho\|\rho') \leq |1 - 2\eta|.$$

因此，

$$D_{e^\varepsilon}(\mu Q^{(\eta)} \| \nu Q^{(\eta)}) \leq \theta \, |1 - 2\eta| \leq |1 - 2\eta| \, \delta,$$

从而得到所声称的隐私参数 $\varepsilon' = \varepsilon$ 和 $\delta' = |1 - 2\eta|\delta$。 □

表 1. 数据集描述。

| 数据集 | 图像尺寸 | 训练集 | 测试集 | 标签数量 | 描述 |
|---|---|---|---|---|---|
| MNIST | 28×28 | 60,000 | 10,000 | 10 | 手写数字 |
| USPS | 16×16 | ≈ 7,300 | ≈ 2,000 | 10 | 扫描的美国邮政信封 |
| FashionMNIST | 28×28 | 60,000 | 10,000 | 10 | 服装物品 |
| CIFAR-10 | 32×32× 3 | 50,000 | 10,000 | 10 | 自然物体 |

## C. 实现

我们使用 Python 3.8 实现所有实验。每项实验均在一台安装了 Linux 64 位操作系统的 GPU 辅助计算节点上进行。我们的测试平台总计配备 72 个 CPU 核心和 377 GB 内存。分配给我们的节点还配备了 2 个 GPU，每个 GPU 的显存为 40GB。

**HYPER-Q 的实现HYPER-Q。** HYPER-Q 使用 PennyLane 量子机器学习模拟器实现（Bergholm 等人，2022年）。具体架构实现了第4节提出并分析的一般机制。具体而言，每个输入图像首先经过两个卷积层，每层后接批归一化和最大池化，以缩小空间维度并提取显著特征。随后，将所得特征图展平，并输入两个全连接层，以生成低维特征向量。接着，将该向量编码到一个量子电路中（包含 5–10 个量子比特）；该电路由三层交替排列的单量子比特旋转（通过 RX 门实现）和纠缠层构成。这对应于编码函数 $f_{\text{enc}}$，其中厄米生成元由RX 门给出。纠缠层采用环形排列的CNOT 门，使每个量子比特 $i$ 与量子比特 $i+1$纠缠，最后一个量子比特则与第一个量子比特纠缠。在计算基上执行投影测量以提取量子输出，然后通过最后一个全连接层处理，得到预测结果。

此外，经典噪声和量子噪声水平设置如下。给定目标差分隐私预算 $(\varepsilon', \delta')$，我们首先固定量子退极化噪声因子 $\eta$，然后校准经典高斯噪声方差 $\sigma^2$，使其根据定理 4.1满足该预算。具体而言，$\sigma^2$ 的选取使经典机制 $A$ 达到 $(\varepsilon, \delta)$-差分隐私，其中

$$\varepsilon = \varepsilon', \qquad \delta = \frac{\delta' - \frac{\eta(1-e^{\varepsilon})}{d}}{1-\eta}.$$

随后，使用解析高斯机制（Balle & Wang，2018年）计算方差 $\sigma^2$，确保经典机制 $A$ 满足 $(\varepsilon, \delta)$-差分隐私，且复合机制 $Q^{(\eta)} \circ A$ 满足目标 $(\varepsilon', \delta')$-差分隐私。

**量子机器学习架构的选择依据。** 我们在实验评估中选择的量子机器学习架构符合量子机器学习和量子差分隐私文献中的既有惯例：受模拟器和当前硬件的计算能力限制，采用 2 到 12 个量子比特的模型是标准做法（Havlíček 等人，2019年；Du 等人，2021a；Hur et al.，2022年；Melo et al.，2023年；Watkins 等人，2023年；Long 等人，2025年）。本研究旨在分析混合量子机器学习模型中 HYPER-Q 差分隐私机制的性能和稳定性，而非优化底层量子机器学习模型以实现最先进的分类准确率。尽管大规模噪声量子训练仍是一项开放挑战，面临贫瘠高原和电路深度约束等问题，但我们通过第 HYPER-Q 节和第 F.4 节的实验结果证明，随着量子比特数量增加，我们的机制能够有效扩展且保持稳定；相关结果见第 F.6节。

## D. 数据集与基线的描述

**数据集：** 我们在三个图像分类数据集上评估我们的方法：MNIST（Lecun 等人，1998）、FashionMNIST（Xiao 等人，2017）和 USPS（Hull，2002）。表 1 分别简要描述了它们。

**经典差分隐私机制：** 为了评估我们的差分隐私机制 HYPER-Q，我们将其与经典差分隐私机制进行比较。下面总结了比较中使用的差分隐私机制：

- **基础高斯。** 高斯机制是实现（$\varepsilon$，$\delta$）-差分隐私的标准方法。高斯机制会从均值为零、方差根据预先设定的隐私参数校准的正态分布中引入噪声（Dong 等人，2022年）。它非常适用于敏感度使用 $\ell_2$ 范数度量的函数。给定一个函数$f$，其构造界为 $L_{\text{cons}}$，且使用 $\ell_2$ 范数度量，该机制通过添加噪声 $\mathcal{N}(0, \sigma^2 I)$实现（$\varepsilon$，$\delta$）-差分隐私

其中 $\sigma$ 的计算方式如下:

$$\sigma = \sqrt{2\ln\left(\frac{1.25}{\delta}\right)} L_{\text{cons}}/\varepsilon$$

• **解析高斯机制**。 解析高斯机制通过利用由隐私损失分布推导出的更紧界限，改进了基础高斯方法（Balle & Wang，2018年）。具体而言，我们可以隐式刻画隐私损失（Cullen 等人，2024年）：

$$\delta(\varepsilon) = \Phi\left(-\frac{L_{\text{cons}}}{2\sigma} + \frac{\varepsilon\sigma}{L_{\text{cons}}}\right) - e^{\varepsilon} \cdot \Phi\left(-\frac{L_{\text{cons}}}{2\sigma} - \frac{\varepsilon\sigma}{L_{\text{cons}}}\right)$$

其中， $\Phi(\cdot)$ 是标准高斯分布的累积分布函数。该表述使我们能够通过数值方法求解满足目标（$\varepsilon$， $\delta$）所需的最小 $\sigma$。

• **DP-SGD**。 差分隐私随机梯度下降（DP-SGD）算法通过裁剪逐样本梯度，并在每个优化步骤中添加经过校准的高斯噪声来实现隐私保护（Abadi 等人，2016年）。作为一种梯度扰动机制，DP-SGD 旨在提供 （$\varepsilon$， $\delta$）-DP，同时保持与大规模深度学习的兼容性。隐私保证源于对梯度更新敏感度的控制，以及注入与该敏感度成比例的噪声。尽管 DP-SGD 已成为经典机器学习中广泛确立的基线方法，但其在量子机器学习环境中的表现仍基本未被探索。除（Watkins 等人，2023年）中介绍的直接应用外，DP-SGD 的梯度噪声与量子梯度估计之间的相互作用尚未得到系统研究。为完整起见，我们将 DP-SGD 纳入比较，作为一种具有代表性的梯度级经典机制。

**经典机器学习模型：**为评估实际可行性，我们将配备 HYPER-Q 的量子机器学习模型与三种配备经典差分隐私机制的经典机器学习模型进行比较：多层感知机（MLP）、ResNet-9 和视觉 Transformer（ViT）。下文描述这些经典机器学习模型的实现:

• **多层感知机：**我们实现了一个多层感知机，其前馈网络由全连接层和 ReLU 激活组成。它包含一个具有 100 个单元的隐藏层，以及一个与类别数相对应的最终线性输出层。它与 Sci-Kit Learn 库中的默认多层感知机相同 [1]。由于该库并非开箱即用地适配图形处理器，我们在不使用该库的情况下实现了它。我们的从零开始 版本可在 GPU 上并行化。

• **ResNet-9：**我们实现了一个受原文启发的 ResNet-9模型（He 等人，2016年）。它由一系列卷积层和两个包含跳跃连接的残差块组成。它通过递增的特征维度处理输入： [32, 64, 128]。我们在整个模型中采用批归一化和 ReLU 激活，随后是最大池化层。该模型以一个用于分类的全连接层结束。

• **ViT：**我们实现了一个受（Dosovitskiy 等人，2021）启发的 ViT 模型。该模型将输入图像划分为互不重叠的图像块，并在线性嵌入后添加位置编码和类别标记。多个自注意力层处理每个序列，然后通过应用于类别标记的全连接头进行分类。

## E. 对抗训练与测试集

我们通过受 PixelDP 机制启发的对抗训练与测试框架，评估 HYPER-Q 的对抗鲁棒性（Lecuyer 等人，2019年）。与 PixelDP 类似，在训练期间，我们定义一个 构造攻击界$L_{cons}$ ，以 $\ell_2$ 范数来表示理论鲁棒性保证。具体而言，该界限确定了允许的最大对抗扰动，在此扰动范围内，模型经认证能够保持其预测能力。在实验中，我们将该值设为 $L_{cons} = \{0.1, 0.2, 0.3, 0.4\}$。对于 HYPER-Q 和经典基线模型，$\ell_2$ 噪声都直接注入输入。尽管量子模型和经典模型的底层架构存在差异，这种设置仍可公平比较二者的鲁棒性保证。

[1] https://scikit-learn.org/stable/

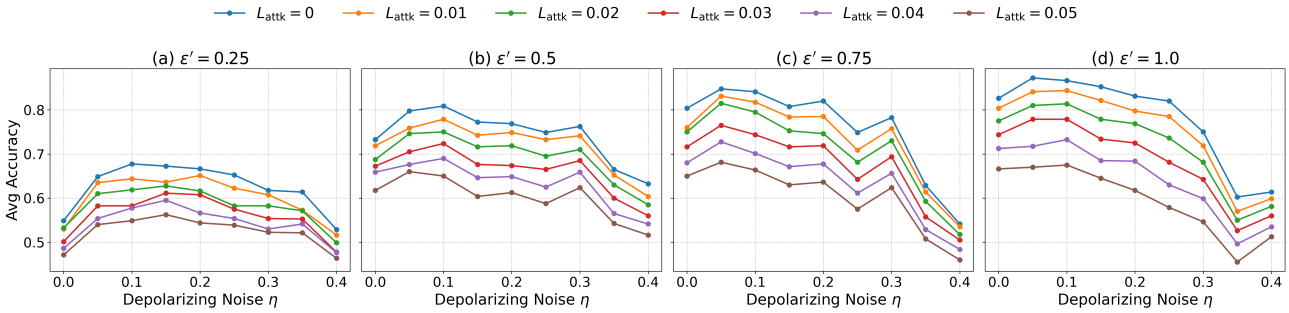



图 *4.* 退极化噪声 $\eta$对 MNIST 数据集模型效用的影响。子图 (a)–(d) 展示了不同隐私预算 $\varepsilon' \in \{0.25, 0.5, 0.75, 1.0\}$下的性能。每条曲线表示模型在不同强度 $L_{\text{attk}} \in \{0, \ldots, 0.05\}$的 FGSM 攻击下的平均准确率。

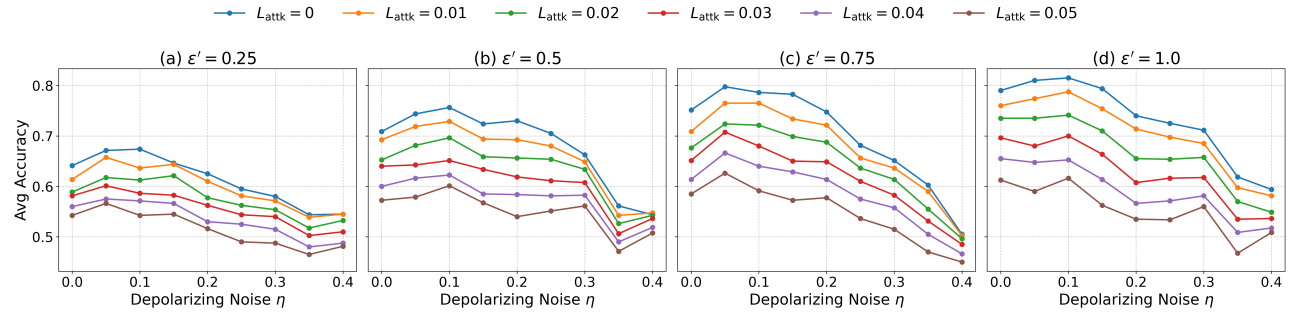



图 *5.* 退极化噪声 $\eta$对 Fashion-MNIST 数据集模型效用的影响。子图 (a)–(d) 展示了不同隐私预算 $\varepsilon' \in \{0.25, 0.5, 0.75, 1.0\}$下的性能。每条曲线表示模型在不同强度 $L_{\text{attk}} \in \{0, \ldots, 0.05\}$的 FGSM 攻击下的平均准确率。

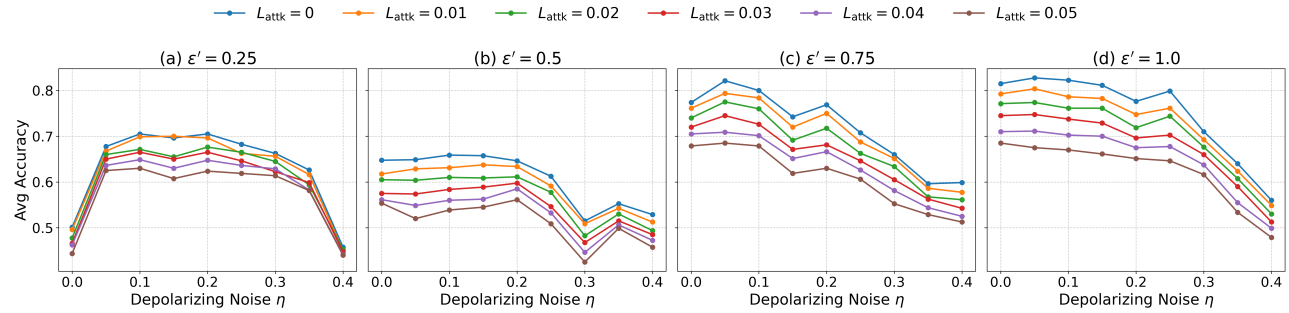



图 *6.* 退极化噪声 $\eta$对 USPS 数据集模型效用的影响。子图 (a)–(d) 展示了不同隐私预算 $\varepsilon' \in \{0.25, 0.5, 0.75, 1.0\}$下的性能。每条曲线表示模型在不同强度 $L_{\text{attk}} \in \{0, \ldots, 0.05\}$的 FGSM 攻击下的平均准确率。

为评估超出认证保证范围的经验鲁棒性，我们针对每个模型测试其在以下范数约束下的对抗扰动： $\ell_\infty$ 范数。具体而言，对于每个 $L_{cons}$ 值，我们使用经验攻击界限 $L_{attk}$进行实验。在实验中，我们将该值设为 $L_{attk} = \{0, 0.01, 0.02, 0.03, 0.04, 0.05\}$，同时实施两种对抗攻击：快速梯度符号法（FGSM）和投影梯度下降（PGD）。借此，我们能够观察模型在真实威胁下的性能；这些威胁可能不符合认证威胁模型的约束。此外，我们采用（Cohen等人，2019年）提出的随机平滑技术，为对抗样本提供认证预测。

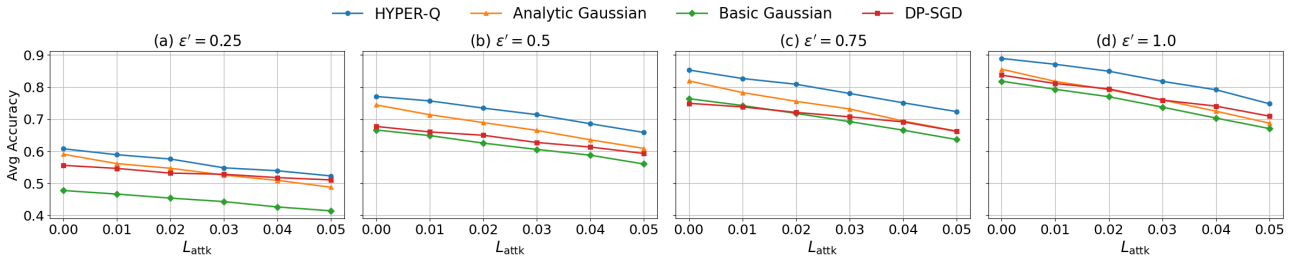



图 7. 在 MNIST数据集上遭受 FGSM 攻击时，不同 $\varepsilon'$ 值和 $\delta' = 1 \times 10^{-5}$下各类加噪机制的准确率。对于每一对 $(L_{\text{attk}},\ \varepsilon')$，报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均所得。使用 $\eta = 0.1$对 HYPER-Q 进行评估。

## F. 补充实验

### F.1. $\eta$ 的敏感性分析（扩展）

本节将进一步介绍第 5.1 节中的敏感性分析。图 4、5 和 6 共同表明，在三个数据集 MNIST、Fashion-MNIST 和 USPS 上，退极化噪声 $\eta$ 与模型性能之间的关系呈现出高度一致的结构。尽管数据集的复杂度各不相同，准确率曲线却具有相同的单峰形状。具体而言，随着 $\eta$ 逐渐偏离 0，性能最初会提高，达到峰值后，随着量子失真占据主导地位而下降。这一模式在每个隐私预算 $\varepsilon' \in \{0.25, 0.5, 0.75, 1\}$ 以及所有攻击界限 $L_{\text{attk}}$下均可观察到。

在比较各数据集性能峰值的位置时，我们观察到高度一致的趋势。在更严格的隐私预算（$\varepsilon' \leq 0.5$）下，最佳性能通常出现在 $\eta = 0.1$ 或 $\eta = 0.05$。例如，MNIST和Fashion-MNIST在 $\eta = 0.1$ 处达到峰值，USPS同样在 $\eta = 0.1 - 0.15$ 处达到峰值（对应于 $\varepsilon' = 0.25$）。随着隐私要求更为宽松（$\varepsilon' \geq 0.75$），三个数据集的峰值均向更小的噪声水平移动，通常为 $\eta = 0.05$。我们观察到，最优 $\eta$ 始终落在狭窄区间0.05-0.15内。尽管 $\eta$ 在实践中往往难以精确校准（Hu等人，2023年），这一稳定的峰值范围提供了稳健的指导，使用户能够可靠地校准 $\eta$于该区间内，而无需进行穷尽式扫描。

### F.2. HYPER-Q的鲁棒性分析（扩展）

与第5.2节相同，我们评估了HYPER-Q 在退极化噪声水平 $\eta = 0.1$下的对抗鲁棒性。我们将其性能与基础高斯、解析高斯和 DP-SGD 机制进行比较，并确保所有方法都在相同的隐私预算下，应用于相同的量子机器学习模型。图7、8 和 9 分别展示了在 FashionMNIST 和 USPS 数据集上进行 FGSM 攻击的结果，其中 $\varepsilon' \in \{0.25, 0.5, 0.75, 1\}$。在 MNIST 和 FashionMNIST 上，HYPER-Q 在所评估的各个隐私预算下，始终优于所有基线机制。相比之下，在 USPS 上，解析高斯在较大的隐私预算下（$\varepsilon' > 0.25$）取得最高准确率，而HYPER-Q则在最严格的隐私约束下（$\varepsilon' = 0.25$）表现最佳。

图10、11和12 分别报告了HYPER-Q 及基线机制在 MNIST、FashionMNIST 和 USPS 上遭受 PGD 攻击时的性能。观察到的趋势与 FGSM 攻击下的趋势十分相似。具体而言，HYPER-Q 在 MNIST 和 FashionMNIST 上始终取得最高准确率，而解析高斯在 USPS 上表现最佳。

### F.3. 与经典模型的比较基准测试

该HYPER-Q 机制专为量子机器学习模型设计。这引出了一个关于实际可行性的关键问题：受 *HYPER-Q* 保护的量子机器学习模型，能否媲美或超越受各自传统隐私机制保护的经典模型？。我们展示了受 HYPER-Q保护的量子机器学习模型（采用经验上最优的量子噪声设置，$\eta = 0.1$）与三种经典基线模型——ResNet-9、ViT 和多层感知机——之间的性能比较；这三种模型均采用解析高斯噪声保护。图 13、14 和15 分别展示了所有模型在 MNIST、FashionMNIST 和 USPS 数据集上遭受 FGSM 攻击时的性能比较。在图 14中，我们观察到，在所有 $\varepsilon'$取值下，ResNet-9模型的表现均优于 HYPER-Q 和其他基线模型。不过值得注意的是，当 HYPER-Q 取值较大时，其性能与 ResNet-9模型非常接近。只有当 $\varepsilon'$取值更高时， $L_{attk}$ 我们才会

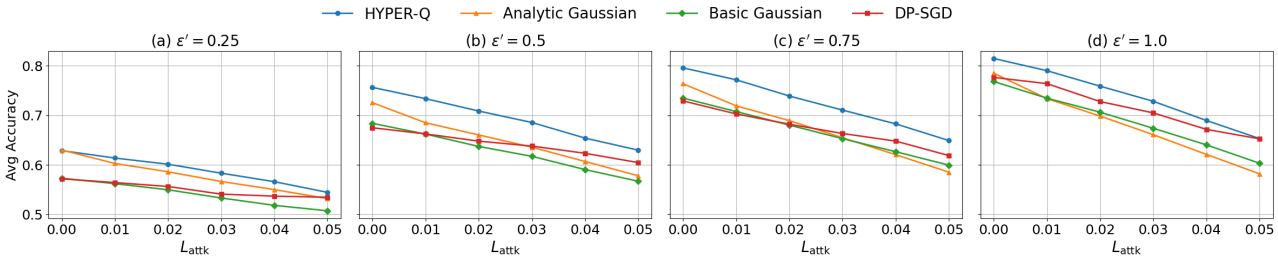



图 *8.* 在 FashionMNIST数据集上遭受 FGSM 攻击时，不同 $\varepsilon'$ 取值和 $\delta' = 1 \times 10^{-5}$下各种加噪机制的准确率。对于每一对 ($L_{\text{attk}}$，$\varepsilon'$)，所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均值所得。使用 $\eta = 0.1$对 HYPER-Q 进行评估。

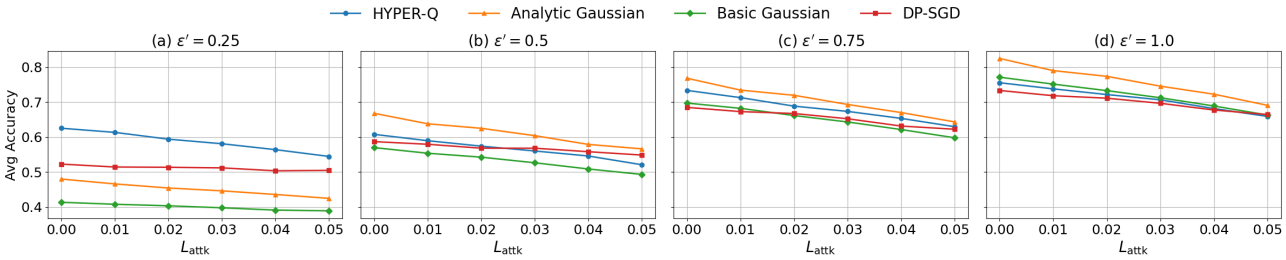



图 *9.* 在 USPS数据集上遭受 FGSM 攻击时，不同 $\varepsilon'$ 取值和 $\delta' = 1 \times 10^{-5}$下各种加噪机制的准确率。对于每一对 ($L_{\text{attk}}$，$\varepsilon'$)，所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均值所得。使用 $\eta = 0.1$对 HYPER-Q 进行评估。

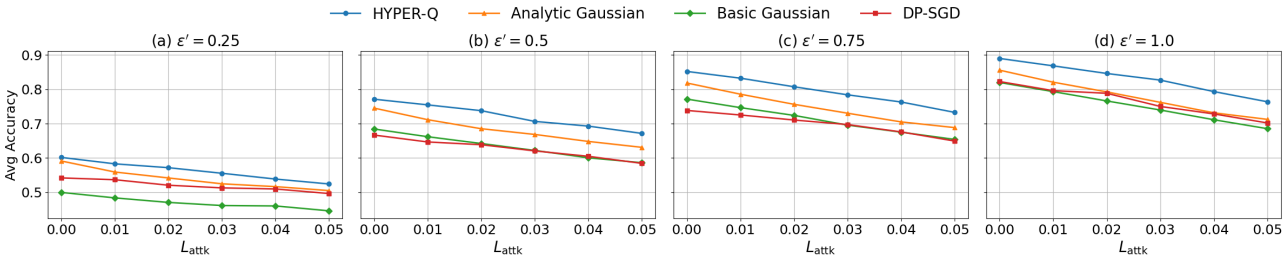



图 *10.* 各种 在 MNIST数据集上遭受 PGD 攻击时，不同 $\varepsilon'$ 取值下的加噪机制 d $\delta' = 1 \times 10^{-5}$。对于每一对 对于 ($L_{\text{attk}}$，$\varepsilon'$)，所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均得到的。HYPER-Q 在 $\eta = 0.$ 下进行检验 1.

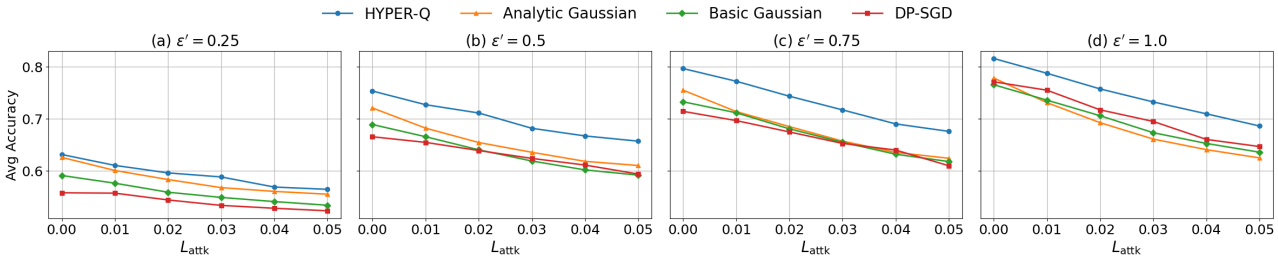



图*11.* FashionMNIST数据集上，各种加噪机制在不同 $\varepsilon'$ 取值和 $\delta' = 1 \times 10^{-5}$ 下遭受 PGD 攻击时的准确率。对于每一对 ($L_{\text{attk}}$，$\varepsilon'$)，所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均得到的。HYPER-Q 在 $\eta = 0.1$ 下进行检验。

观察到这两个模型之间出现明显差距。相反，在 USPS 数据集上，当HYPER-Q 时，其表现优于所有其他基线模型。具体而言，与 ResNet-9模型相比，当 $\varepsilon' \in \{0.25, 0.5\}$时，HYPER-Q 的平均准确率高出 $\approx 30\%$。当 $\varepsilon' = 0.25$ 时，这一数值降至 $\approx 2\%$ 。随着 $\varepsilon' = 0.5$，ResNet-9模型的竞争力有所提升；当 $\varepsilon' \in \{0.75, 1.0\}$时，其表现与 HYPER-Q 相当，而后者分别领先 $\approx 5\%$。一个有趣的现象是，多层感知机的表现略有波动，并且在所有 $\varepsilon'$取值下都迅速退化。HYPER-Q 和其他基线模型在所有取值下则稳定得多。图 16、17和18 展示了对比

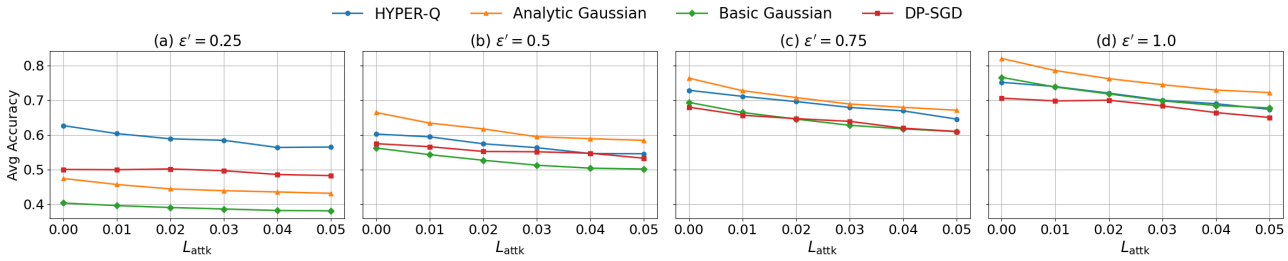



图 *12.* 在不同 $\varepsilon'$ 取值和$\delta' = 1 \times 10^{-5}$下，各种加噪机制在USPS数据集上遭受PGD 攻击时的准确率。对于每一对（$L_{\text{attk}}$，$\varepsilon'$），所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均得到的。HYPER-Q在 $\eta = 0.1$条件下进行考察。

HYPER-Q与我们的基线模型在遭受PGD 攻击时的性能，在本质上与全部三个数据集上FGSM 攻击的结果几乎完全相同。

### F.4. 维度可扩展性的实证分析

为应对HYPER-Q的实际可扩展性问题，我们实证分析了希尔伯特空间维度$d = 2^n$ 对所需经典噪声降低的影响。尽管定理 1引入了一个加性项 $^{\eta}(1-e^{\epsilon})\,d$，该加性项表面上会随系统规模扩大而缩小，但我们的分析表明，隐私放大会趋于稳定，而不会消失。

图19展示了经典噪声方差的平均百分比降幅（$\sigma^2$）随量子比特数量$n$（从1到29）在不同量子噪声水平 $\eta \in [0.05, 0.4]$下的变化。结果突出了两点观察。对于小规模系统，我们观察到所需经典噪声大幅降低，超过90%（当 $n < 3$ 时）。在该区间内，维度相关项 $\frac{1}{d}$ 在定理 1 中占主导地位，为隐私预算提供了显著增益。另一方面，随着量子比特数量增加且 $1/d$ 项消失，噪声降幅并未降至零。相反，曲线趋于平稳，稳定在非零水平。这代表了与规模无关的乘性放大（$1 - \eta$），该结果由我们的理论框架推导得出。例如，在 $\eta = 0.4$下，该机制保持约为8% 的稳定噪声降幅，即使在 $n = 29$ （其中 $d \approx 5 \times 10^8$）时亦然。

### F.5. 效用界紧致性的实证验证

为严格评估定理 3 中推导的理论效用界的紧致性，我们进行了实证分析，将观测到的最坏情况误差与该界进行比较。定理 3 通过界定最大偏差 $\text{Error} = \sup_x |\mathcal{M}_{\text{full}}(x) - \mathcal{M}_{\text{clean}}(x)|$，刻画了该机制的稳定性。该界表明：

$$\Pr\left(\text{Error} \le L_\infty \cdot \sigma\sqrt{2\ln\frac{2d_X}{p}} + 2\eta\|E_{\text{exp}}\|_{op}\right) \ge 1 - p$$

其中 $L_\infty = 2(1-\eta)\|E_{\text{exp}}\|_{op}\|W\|_\infty(\sum_j \|H_j\|_{op})$。

我们通过测量模拟中观测到的最大经验误差与理论界 $\text{Bound}(\sigma,\ \eta,\ p) = L_\infty\sigma\sqrt{2\ln(2d_X/p)} + 2\eta||E_{\text{exp}}||_{\text{op}}$ 之间的比率，评估该界的紧致性。具体而言，给定样本集 $S$，该比率计算如下：

$$\text{Ratio} = \frac{\max_{x\in S}|\mathcal{M}_{\text{full}}(x) - \mathcal{M}_{\text{clean}}(x)|}{\text{Bound}(\sigma,\eta,p)}$$

在本实验中，我们将失败概率设为 $p = 0.01$。这确保理论界以 99% 的置信水平成立。我们使用样本量 $|S| = 10000$，针对每种（$\sigma,\eta$）配置计算了比率；比率越接近 1，表明界越紧。图20展示了参数网格中各项比率的结果。我们观察到，在低噪声设置下，例如（$\sigma,\eta$）$=$(0.5,0)，比率达到较大的数值（如 0.923），证实该界能够有效刻画最坏情况误差。尽管随着总噪声幅度增大，该界会变得较松，但仍然具有实质意义。值得注意的是，在不存在经典噪声（$\sigma = 0$）时，所有 $\eta$ 的比率都保持不变。这一现象归因于退极化信道的线性性：经验误差和理论量子项（$2\eta\|E_{\text{exp}}\|_{\text{op}}$）均随 $\eta$线性变化。

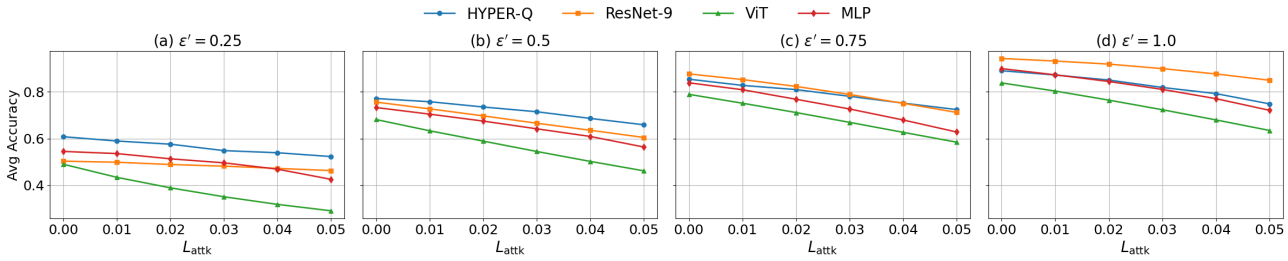



图 *13.* 在 MNIST数据集上遭受 FGSM 攻击时，采用 HYPER-Q 保护的量子机器学习模型与采用解析高斯噪声保护的三个经典基线（ResNet-9、ViT 和多层感知机）的准确率对比。HYPER-Q 模型采用经实证确定的最佳量子噪声设置 ($\eta = 0.1$) 进行评估。对于每个 $(L_{\text{attk}}, \varepsilon')$ 参数对，报告的准确率是对所有 $L_{\text{cons}}$ 种设置取平均所得。$\delta' = 1 \times 10^{-5}$ 适用于所有设置。

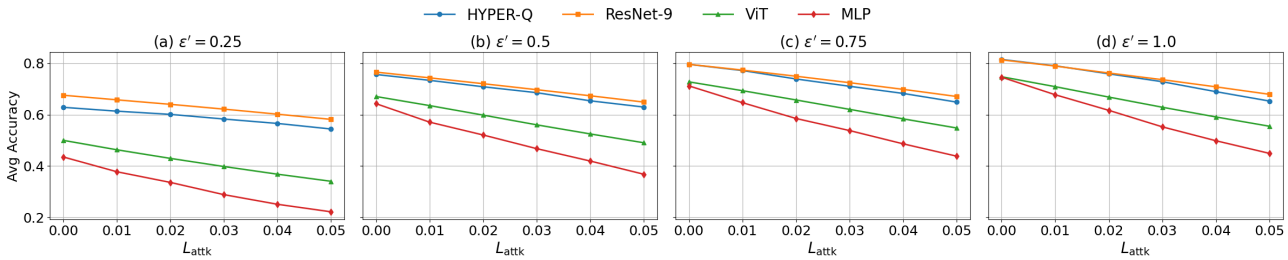



图 *14.* 在 FashionMNIST数据集上遭受 FGSM 攻击时，采用 HYPER-Q 保护的量子机器学习模型与采用解析高斯噪声保护的三个经典基线（ResNet-9、ViT 和多层感知机）的准确率对比。HYPER-Q 模型采用经实证确定的最佳量子噪声设置 ($\eta = 0.1$) 进行评估。对于每个 $(L_{\text{attk}}, \varepsilon')$ 参数对，报告的准确率是对所有 $L_{\text{cons}}$ 种设置取平均所得。$\delta' = 1 \times 10^{-5}$ 适用于所有设置。

### F.6. CIFAR-10 上的性能分析

为评估 HYPER-Q 在更复杂数据上的鲁棒性，我们将实验扩展至 CIFAR-10。与 MNIST 和 USPS 相比，CIFAR-10 是一个难度显著更高的基准数据集。不同于灰度数据集，CIFAR-10 由 RGB图像构成，具有更高的多样性和更丰富的特征结构，因此需要更大的量子特征映射。在这一设置下，我们采用一个 10 量子比特的变分量子机器学习模型，并在固定隐私预算 $\varepsilon' = 1$ 下将 HYPER-Q 与三个经典基线进行比较。

图 21 展示了测试准确率随攻击强度 $L_{\text{attk}} \in \{0, 0.01, \ldots, 0.05\}$ 的变化情况。所有模型的准确率都会随着攻击强度增加而下降，但下降速度差异显著。HYPER-Q 的初始准确率为 73.9%（$L_{\text{attk}} = 0$），随后平稳下降至 47.7%（$L_{\text{attk}} = 0.05$）。这一退化趋势与 ResNet-9 相当：ResNet-9 的基线准确率较高，为 86.0%，但在最高攻击上界下同样降至 48.7%。相比之下，ViT 和多层感知机的性能下降要快得多，准确率分别从初始的 75.9% 和 69.3% 降至 $L_{\text{attk}} = 0.05$ 时的仅 14.2% 和 10.8%。

这些结果揭示了两点。首先，即使面对需要更深层量子表征的高维图像数据集，HYPER-Q 在中等攻击强度下仍能与经典基线相媲美。其次，尽管经典深度模型具有更高的无扰动准确率，但随着扰动增强，其鲁棒性会急剧下降；相比之下，HYPER-Q 的准确率下降则更为平稳、可控。这表明，在更复杂、量子比特数更高的量子机器学习设置中，HYPER-Q 的混合噪声机制仍能带来显著的效用优势。

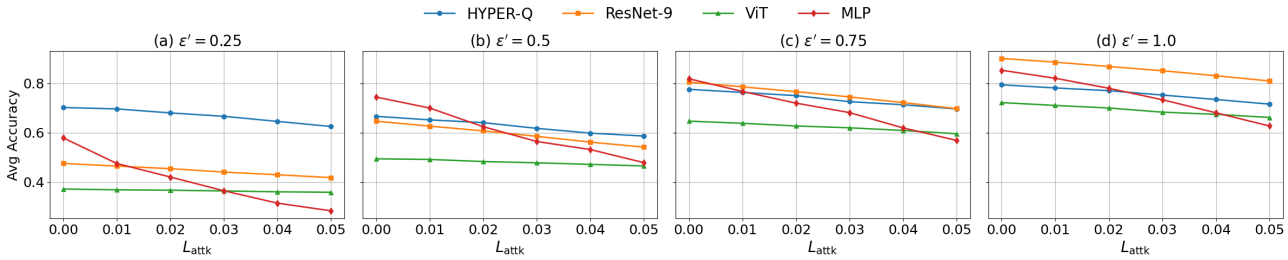



图*15.* 在 USPS数据集上的 FGSM 攻击下，受 HYPER-Q 保护的量子机器学习模型与受解析高斯噪声保护的三个经典基线（ResNet-9、ViT 和多层感知机）之间的准确率比较。HYPER-Q 模型采用其经验上最佳的量子噪声设置（$\eta = 0.1$）进行评估。对于每个（$L_{\text{attk}}, \varepsilon'$）对，所报告的准确率是对所有 $L_{\text{cons}}$ 设置取平均得到的。 $\delta' = 1 \times 10^{-5}$ 适用于所有设置。

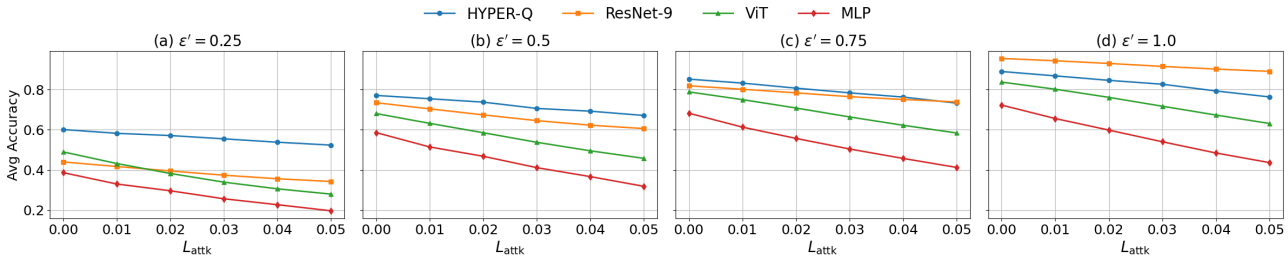



图 *16.* 在针对 MNIST数据集的 PGD 攻击下，受 HYPER-Q 保护的量子机器学习模型与受解析高斯噪声保护的三个经典基线（残差网络-9、ViT 和多层感知机）的准确率对比。HYPER-Q 模型采用其经验上表现最佳的量子噪声设置（$\eta = 0.1$）进行评估。对于每个（$L_{\text{attk}}$， $\varepsilon'$）组合，报告的准确率为所有 $L_{\text{cons}}$ 设置的平均值。 $\delta' = 1 \times 10^{-5}$ 适用于所有设置。

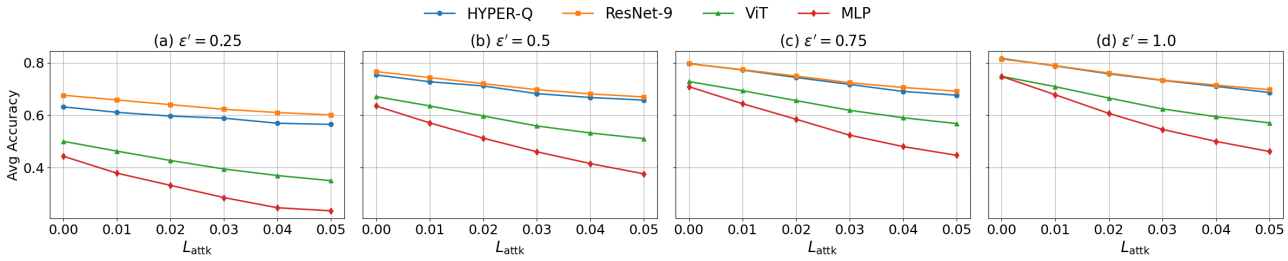



图 *17*。 在 FashionMNIST数据集上进行 PGD 攻击时，采用 HYPER-Q 保护的 QML 模型与采用解析高斯噪声保护的三个经典基线（ResNet-9、ViT 和多层感知机）的准确率比较。HYPER-Q 模型采用经验上表现最佳的量子噪声设置（$\eta = 0.1$）进行评估。对于每个（$L_{\text{attk}}$、 $\varepsilon'$）组合，所报告的准确率均取所有 $L_{\text{cons}}$ 设置下的平均值。$\delta' = 1 \times 10^{-5}$ 对于所有设置。

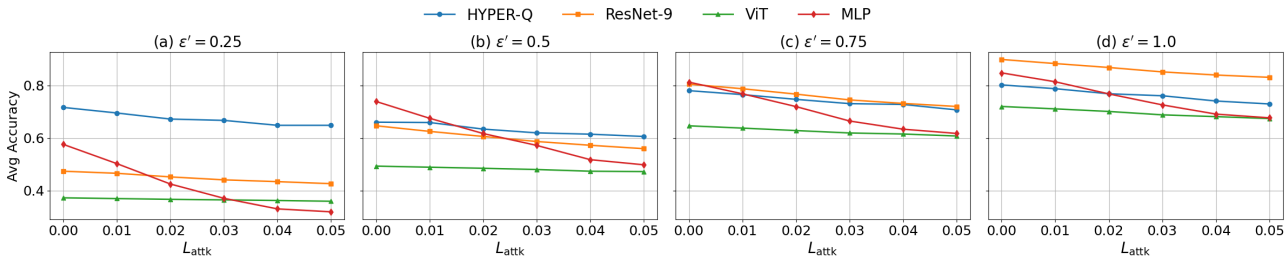



图 *18*。 在 USPS数据集上进行 PGD 攻击时，采用 HYPER-Q 保护的 QML 模型与采用解析高斯噪声保护的三个经典基线（ResNet-9、ViT 和多层感知机）的准确率比较。HYPER-Q 模型采用经验上表现最佳的量子噪声设置（$\eta = 0.1$）进行评估。对于每个（$L_{\text{attk}}$、 $\varepsilon'$）组合，所报告的准确率均取所有 $L_{\text{cons}}$ 设置下的平均值。 $\delta' = 1 \times 10^{-5}$ 对于所有设置。

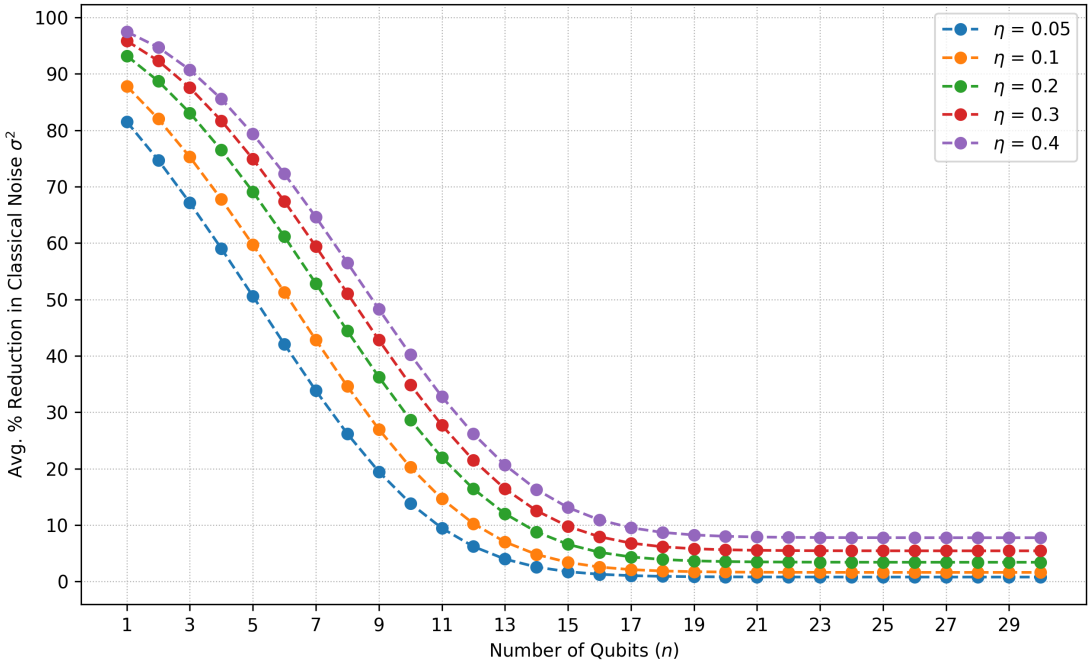
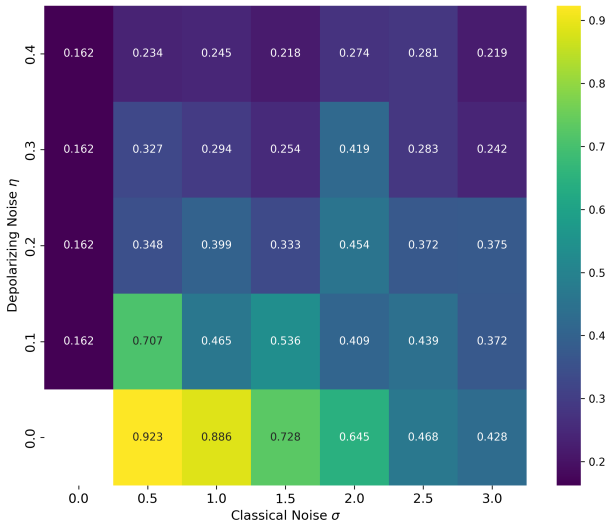



图 *19.* 量子噪声水平对不同量子比特扩展规模下经典噪声降低的影响




*Fi*图 *20.* 不同 ($\eta$, $\sigma$) 值下实际效用损失与理论界之比的热力图 s.

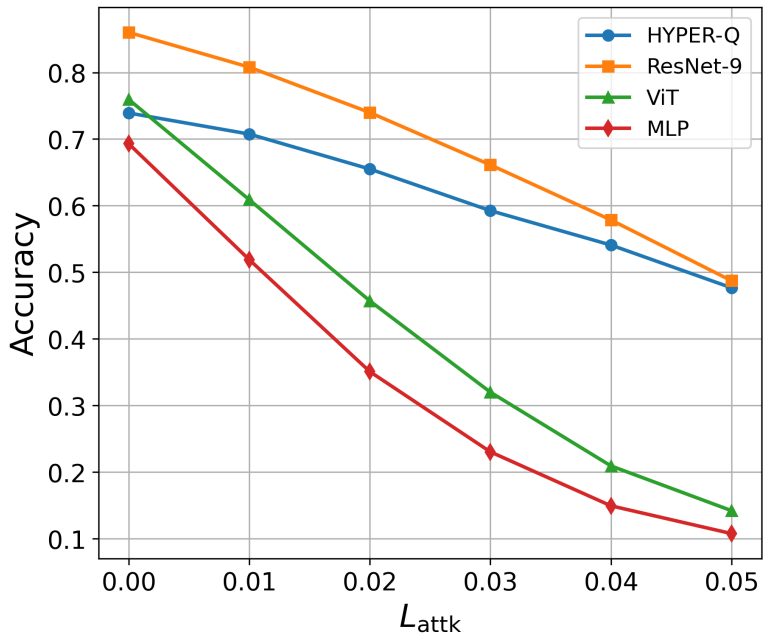



图 *21.* 在隐私预算固定为 $\varepsilon' = 1$ 时，HYPER-Q 与经典基线（ResNet-9、ViT 和多层感知机）在 CIFAR-10 上面对不同攻击强度 $L_{\text{attk}}$ 时的性能。