

# Communication-Efficient and Privacy-Adaptable Mechanism – a Federated Learning Scheme with Convergence Analysis

Chun Hei Michael Shiu

Department of Electrical and Computer Engineering, University of British Columbia  
 shiuchm@ece.ubc.ca

Chih Wei Ling

Department of Computer Science, City University of Hong Kong  
 cwling6@um.cityu.edu.hk

## Abstract

Federated learning enables multiple parties to jointly train learning models without sharing their own underlying data, offering a practical pathway to privacy-preserving collaboration under data-governance constraints. Continued study of federated learning is essential to address key challenges in it, including communication efficiency and privacy protection between parties. A recent line of work introduced a novel approach called the Communication-Efficient and Privacy-Adaptable Mechanism (CEPAM), which achieves both objectives simultaneously. CEPAM leverages the rejection-sampled universal quantizer (RSUQ), a randomized vector quantizer whose quantization error is equivalent to a prescribed noise, which can be tuned to customize privacy protection between parties. In this work, we theoretically analyze the privacy guarantees and convergence properties of CEPAM. Moreover, we assess CEPAM’s utility performance through experimental evaluations, including convergence profiles compared with other baselines, and accuracy-privacy trade-offs between different parties.

## Index Terms

federated learning, randomized quantization, differential privacy, convergence analysis, channel simulation

## I. INTRODUCTION AND MOTIVATION

Federated learning (FL) was first introduced in [1] as a framework where many decentralized clients collaboratively train a shared global model by sending only updates (e.g. gradients, model differences) to a central server, without transmitting data from their underlying databases. Such decentralized design [2] exploits their computational resource to handle vast amounts of data held by each client. To make such framework practical, the authors in [1] propose the FederatedAveraging (FedAvg) approach, empirically showing that it can train deep neural networks efficiently. Unlike in classical centralized machine learning, communication efficiency and privacy protection are primary concerns [3] in the study of federated learning, due to the distributed nature and the need to limit information exchange between parties. Traditionally, these challenges have been addressed separately, with compression or quantization schemes [4]–[8] improving communication efficiency, and differential privacy (DP) mechanisms [9], [10] enhancing privacy protection.

To address communication efficiency and privacy protection simultaneously, a common strategy is to first apply a differential privacy (DP) mechanism (e.g. Gaussian [10] or Laplace [9] noise), and then quantize the perturbed updates [11], [12]. However, this DP-then-quantize pipeline is suboptimal: as the aggregated error from the privacy mechanism and quantization scheme is no longer exactly controlled, which harms the model accuracy. Also, the quantization error itself does not contribute to privacy protection against other curious clients. An alternative line of work uses randomized quantization such as subtractive dithering [13]–[17], to jointly achieve compression and privacy. However, methods based on subtractive dithering [18]–[22] typically rely on scalar quantization, which is less efficient than vector quantization. In addition, they often support only certain noise distributions such as Laplace or multivariate- $t$  distributions [23],<sup>1</sup> rather than fully encompassing the Gaussian mechanism.

Another related line of work is the study of channel simulation [24]–[31], where one seeks to reproduce the effect of a noise channel with or without shared randomness. Based on the observation that Gaussian noise can be rewritten as a mixture of uniform distributions [32], [33], constructions in [34], [35] combine randomized quantization and dithering to simulate any one-dimensional additive-noise channel, and more recent results [36] extend these ideas to general multivariate distributions via randomized vector quantization. In particular, the rejection-sampled universal quantizer (RSUQ) [36] has been shown to exactly simulate any multivariate non-uniform continuous additive-noise channel with finite communication.

More recently, a joint mechanism called the Communication-Efficient and Privacy-Adaptable Mechanism (CEPAM) has been proposed [37] for the FL setting. CEPAM is built upon RSUQ by exploiting its ability to reinterpret quantization distortion as

<sup>1</sup>Although [23] proposes adding a low-power privacy-preserving noise before quantization, it does not specify how to design the necessary infinitesimal noise for local DP’s Gaussian noise, a highly non-trivial task. In contrast, our CEPAM approach provides a clear specification for generating Gaussian noise under the central DP model.

an additive noise term with controllable variance, independent of the underlying model updates. In particular, by appropriately configuring RSUQ,<sup>2</sup> CEPAM can emulate different privacy mechanisms while simultaneously compressing the updates, thereby enabling a tunable trade-off between privacy protection and communication cost within a unified framework.

This work is a continuation of the study of CEPAM in the FL setting. In contrast to [37], which studies the performance of CEPAM when compressing the model differences as updates, we consider a slightly modified setup in which CEPAM is applied to compress the stochastic gradients. Moreover, whereas the earlier work reported empirical convergence behavior of CEPAM without a formal analysis, we provide a rigorous convergence analysis together with theoretical privacy guarantees for the proposed scheme.

We theoretically analyze the convergence behavior of CEPAM, and validate our convergence results for both CEPAM-Gaussian and CEPAM-Laplace through experimental evaluations. Experimental results demonstrate that both CEPAM-Gaussian and CEPAM-Laplace achieve improvements of 0.8-1.1% in test accuracy, respectively, compared to several other commonly used baselines in the FL setup. In addition, we investigate the accuracy-privacy trade-offs among clients through experimental evaluations, and the observed behavior coincides with that obtained when compressing model differences in the previous study [37].

The rest of this paper is structured as follows. In Section II, we define the system model and review preliminaries. In Section III, we detail the proposed scheme CEPAM and highlight the modifications. In Section IV, we provide theoretical analysis on the performance of CEPAM, and the corresponding experimental evaluations are given in V. Finally, we conclude the paper in Section VI.

## II. SYSTEM MODEL AND PRELIMINARIES

In this section, we describe the system model and setup used in this paper. We first present the FL framework in Section II-A, followed by a detailed description of LRSUQ in Section II-B. Section II-C reviews the necessary background on differential privacy. Finally, we identify and summarize the desired goals under our framework in Section II-D.

### Notations

Write  $H(X)$  for the entropy and in bits. Logarithms are to the base 2. For  $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^n$ ,  $\beta \in \mathbb{R}$ ,  $\mathbf{G} \in \mathbb{R}^{n \times n}$ ,  $\mathbf{x} \in \mathbb{R}^n$  write  $\beta\mathcal{A} := \{\beta\mathbf{z} : \mathbf{z} \in \mathcal{A}\}$ ,  $\mathbf{G}\mathcal{A} := \{\mathbf{G}\mathbf{z} : \mathbf{z} \in \mathcal{A}\}$ ,  $\mathcal{A} + \mathbf{x} := \{\mathbf{z} + \mathbf{x} : \mathbf{z} \in \mathcal{A}\}$ ,  $\mathcal{A} + \mathcal{B} = \{\mathbf{y} + \mathbf{z} : \mathbf{y} \in \mathcal{A}, \mathbf{z} \in \mathcal{B}\}$  for the Minkowski sum,  $\mathcal{A} - \mathcal{B} = \{\mathbf{y} - \mathbf{z} : \mathbf{y} \in \mathcal{A}, \mathbf{z} \in \mathcal{B}\}$ , and  $\mu(\mathcal{A})$  for the Lebesgue measure of  $\mathcal{A}$ . Let  $B^n := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| \leq 1\}$  be the unit  $n$ -ball. For a function  $f : \mathbb{R}^n \rightarrow \mathbb{R}$ , its superlevel set is defined as  $L_u^+(f) := \{\mathbf{x} \in \mathbb{R}^n : f(\mathbf{x}) \geq u\}$ .

### A. Federated Learning

For the setup, we consider the FL framework (or *federated optimization*) proposed in [1]. More explicitly,  $K$  clients (or devices) possess local dataset  $\mathcal{D}^{(k)}$  where  $k \in \{1, 2, \dots, K\} =: \mathcal{K}$ , work jointly together to train a shared global model  $\mathbf{W}$  with  $m$  parameters through a central server. The goal is to minimize the objective function  $F : \mathbb{R}^m \rightarrow \mathbb{R}$ :

$$\min_{\mathbf{W} \in \mathbb{R}^m} \left\{ F(\mathbf{W}) := \sum_{k \in \mathcal{K}} p_k F_k(\mathbf{W}) \right\}, \quad (1)$$

where  $p_k$  is the weight of client  $k$  such that  $p_k \geq 0$  and  $\sum_{k \in \mathcal{K}} p_k = 1$ . Suppose that the  $k$ -th local dataset contains  $n_k$  training data:  $\mathcal{D}^{(k)} = \{\xi_{k,1}, \dots, \xi_{k,n_k}\}$ . The local objective function  $F_k : \mathbb{R}^m \rightarrow \mathbb{R}$  is defined by

$$F_k(\mathbf{W}) \equiv F_k(\mathbf{W}, \mathcal{D}^{(k)}) := \frac{1}{n_k} \sum_{j=1}^{n_k} \ell(\mathbf{W}; \xi_{k,j}), \quad (2)$$

where  $\ell(\cdot; \cdot)$  is an application-specified loss function.

Let  $T$  denote the total number of iterations in FL and define the set of *synchronization indices*  $\mathcal{T}_T := \{0, \tau, 2\tau, \dots, T\}$ , i.e., the set of integer multiples of some positive integer  $\tau \in \mathbb{Z}^+$  where  $T \equiv 0 \pmod{\tau}$ . We describe one FL round of the FedAvg [1] with some modifications to solve the optimization problem (1) as follows. Let  $\mathbf{W}_t$  denote the global parameter vector available on the server at the time instance  $t \in \mathcal{T}_T$ . At the beginning of each FL round, the server broadcasts  $\mathbf{W}_t$  to all clients. Then, each client  $k$  sets  $\mathbf{W}_t^k = \mathbf{W}_t$  and computes the  $\tau - 1$  ( $\geq 1$ ) local parameter vectors by SGD:<sup>3</sup>

$$\mathbf{W}_{t+t'}^k \leftarrow \mathbf{W}_{t+t'-1}^k - \eta_{t+t'-1} \nabla F_k^{j_t^k}(\mathbf{W}_{t+t'-1}^k), \quad (3)$$

for  $t' = 1, \dots, \tau - 1$ , where  $\eta_{t+t'-1}$  is the learning rate,  $\nabla F_k^j(\mathbf{W}) := \nabla F_k(\mathbf{W}; \xi_{k,j})$  is the gradient computed at a single sample of index  $j$ , and  $j_t^k$  is the sample index chosen uniformly from the local data  $\mathcal{D}^{(k)}$  of client  $k$  at time  $t$ .<sup>4</sup> Finally, client

<sup>2</sup>A variation of RSUQ – layered rejection-sampled universal quantizer (LRSUQ) was used.

<sup>3</sup>When  $T$  is fixed, the larger  $\tau$  is, the fewer the communication rounds there are.

<sup>4</sup>In this work, our focus is on analyzing the computation of a single stochastic gradient at each client during every time instance. The FL convergence rates can potentially be enhanced by incorporating mini-batching technique [38], leaving the detailed analysis for future work.

$k$  computes  $\mathbf{X}_{t+\tau-1}^k = \nabla F_k^{j_{t+\tau}}(\mathbf{W}_{t+\tau-1}^k)$ . Let  $M$  be the maximum  $\ell_2$ -norm of all possible gradients for any given weight vector  $\mathbf{W}$  and sampled dataset  $\xi_k$ , i.e.,

$$M := \sup_{\mathbf{W} \in \mathbb{R}^m, \xi_k \in \mathcal{D}^{(k)}} \mathbb{E} [\|\nabla F_k(\mathbf{W}; \xi_{k,j})\|_2].$$

For simplicity, we assume that all  $K$  clients participate in each FL round for all  $k \in \mathcal{K}$ . The server then aggregates  $K$  local gradients  $\{\mathbf{X}_{t+\tau-1}^k\}_{k \in \mathcal{K}}$ , computes the new global parameter vector:

$$\mathbf{W}_{t+\tau} \leftarrow \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \mathbf{X}_{t+\tau-1}^k, \quad (4)$$

and broadcasts  $\mathbf{W}_{t+\tau}$  to all clients.

The dataset  $\mathcal{D}^{(k)}$  inherently induces a distribution. By an abuse of notation, we also denote this induced distribution as  $\mathcal{D}^{(k)}$ . Suppose the data in client  $k$  is iid sampled from the induced distribution  $\mathcal{D}^{(k)}$ . Thus, the overall distribution becomes a mixture of all local distributions:  $\mathcal{D} = \sum_{k \in \mathcal{K}} p_k \mathcal{D}^{(k)}$ . Previous works typically assume that the data is iid generated by or partitioned among the  $K$  clients, i.e., for all  $k \in \mathcal{K}$ ,  $\mathcal{D}^{(k)} = \mathcal{D}$ . In contrast, we consider a scenario where the data is non-iid (or heterogeneous), implying that  $F_k$  could potentially be an arbitrarily poor approximation to  $F$ .

### B. Rejection-Sampled Universal Quantizer

We briefly review RSUQ [36], which is constructed based on the subtractive dithered quantizer (SDQ) [13], [39], [40].

Given a non-singular generator matrix  $\mathbf{G} \in \mathbb{R}^{n \times n}$ , a *lattice* is the set  $\mathbf{G}\mathbb{Z}^n = \{\mathbf{G}\mathbf{j} : \mathbf{j} \in \mathbb{Z}^n\}$ . A bounded set  $\mathcal{P} \subseteq \mathbb{R}^n$  is called a *basic cell* of the lattice  $\mathbf{G}\mathbb{Z}^n$  if  $(\mathcal{P} + \mathbf{G}\mathbf{j})_{\mathbf{j} \in \mathbb{Z}^n}$  forms a partition of  $\mathbb{R}^n$  [41], [42]. Specifically, the Voronoi cell  $\mathcal{V} := \{\mathbf{x} \in \mathbb{R}^n : \arg \min_{\mathbf{j} \in \mathbb{Z}^n} \|\mathbf{x} - \mathbf{G}\mathbf{j}\| = \mathbf{0}\}$  is a basic cell. Given a basic cell  $\mathcal{P}$ , we can define a *lattice quantizer*  $Q_{\mathcal{P}} : \mathbb{R}^n \rightarrow \mathbf{G}\mathbb{Z}^n$  such that  $Q_{\mathcal{P}}(\mathbf{x}) = \mathbf{y}$  where  $\mathbf{y} \in \mathbf{G}\mathbb{Z}^n$  is the unique lattice point that satisfies  $\mathbf{x} \in -\mathcal{P} + \mathbf{y}$ . The resulting quantization error  $\mathbf{z} := Q_{\mathcal{P}}(\mathbf{x}) - \mathbf{x}$  depends deterministically on the input  $\mathbf{x}$  and is approximately uniformly distributed over the basic cell of the lattice quantizer under some regularity assumptions. Therefore, it is often combined with probabilistic methods such as random dithering to construct SDQ [39], [40].

**Definition 1.** Given a basic cell  $\mathcal{P}$  and a random dither  $\mathbf{v} \sim \text{Unif}(\mathcal{P})$ , a *subtractive dithered quantizer* (SDQ)  $Q_{\mathcal{P}}^{SDQ} : \mathbb{R}^n \times \mathcal{P} \rightarrow \mathbb{R}^n$  for an input  $\mathbf{x} \in \mathbb{R}^n$  is given by  $Q_{\mathcal{P}}^{SDQ}(\mathbf{x}, \mathbf{v}) = Q_{\mathcal{P}}(\mathbf{x} - \mathbf{v}) + \mathbf{v}$ , where  $Q_{\mathcal{P}}$  is the lattice quantizer.

It is well-known that the resulting quantization error of SDQ is uniformly distributed over the basic cell of the quantizer and is statistically independent of the input signal [14], [40], [43]–[45]. However, it may be desirable to have the quantization error follow a uniform distribution over an arbitrary set, rather than being distributed uniformly over a basic cell. RSUQ is a randomized quantizer where the quantization error is uniformly distributed over a set  $\mathcal{A}$ , a subset of a basic cell. This quantization scheme is based on applying rejection sampling on top of SDQ. Intuitively, we keep generating new dither signals until the quantization error falls in  $\mathcal{A}$ .<sup>5</sup>

**Definition 2.** [36, Definition 3] Given a basic cell  $\mathcal{P}$  of the lattice  $\mathbf{G}\mathbb{Z}^n$ , a subset  $\mathcal{A} \subseteq \mathcal{P}$ , and a sequence  $S = (\mathbf{V}_i)_{i \in \mathbb{N}^+}$ ,  $\mathbf{V}_1, \mathbf{V}_2, \dots \stackrel{iid}{\sim} \text{Unif}(\mathcal{P})$  are i.i.d. dithers, the *rejection-sampled universal quantizer* (RSUQ)  $Q_{\mathcal{A}, \mathcal{P}} : \mathbb{R}^n \times \prod_{i \in \mathbb{N}^+} \mathcal{P}_i \rightarrow \mathbb{R}^n$  for  $\mathcal{A}$  against  $\mathcal{P}$  is given by

$$Q_{\mathcal{A}, \mathcal{P}}(\mathbf{x}, (\mathbf{v}_i)_i) := Q_{\mathcal{P}}(\mathbf{x} - \mathbf{v}_h) + \mathbf{v}_h, \quad (5)$$

where

$$h := \min \{i : Q_{\mathcal{P}}(\mathbf{x} - \mathbf{v}_i) + \mathbf{v}_i - \mathbf{x} \in \mathcal{A}\}, \quad (6)$$

and  $Q_{\mathcal{P}}$  is the lattice quantizer for basic cell  $\mathcal{P}$ .

Note that when  $\mathcal{A} = \mathcal{P}$ , SDQ is a special case of RSUQ. RSUQ can be generalized to simulate an additive noise channel with noise following a continuous distribution by using layered construction as in [35], [46]. Consider a pdf  $f : \mathbb{R}^n \rightarrow [0, \infty)$  and write its superlevel set as

$$L_u^+(f) = \{\mathbf{z} \in \mathbb{R}^n : f(\mathbf{z}) \geq u\}.$$

Let  $f_U(u) := \mu(L_u^+(f))$  for  $u > 0$ , which is also a pdf. If we generate  $U \sim f_U$ , and then  $\mathbf{Z}|_{\{U=u\}} \sim \text{Unif}(L_u^+(f))$ , and  $\mathbf{Z} \sim f$  by the fundamental theorem of simulation [47].

**Definition 3.** [36, Definition 4] Given a basic cell  $\mathcal{P}$  of the lattice  $\mathbf{G}\mathbb{Z}^n$ , a probability density function  $f : \mathbb{R}^n \rightarrow [0, \infty)$  where  $L_u^+(f)$  is always bounded for  $u > 0$ , and  $\beta : (0, \infty) \rightarrow [0, \infty)$  satisfying  $L_u^+(f) \subseteq \beta(u)\mathcal{P}$  for  $u > 0$ , and a random pair  $S = (U, (\mathbf{V}_i)_{i \in \mathbb{N}^+})$  where the latent variable  $U \sim f_U$  with  $f_U(u) := \mu(L_u^+(f))$ , and  $\mathbf{V}_1, \mathbf{V}_2, \dots \stackrel{iid}{\sim} \text{Unif}(\mathcal{P})$  is a sequence

<sup>5</sup>It is easy to see that the acceptance probability is  $\mu(\mathcal{A})/\mu(\mathcal{P})$ .

of i.i.d. dither signals, the *layered rejection-sampled universal quantizer* (LRSUQ)  $Q_{f,\mathcal{P}} : \mathbb{R}^n \times \mathbb{R} \times \prod_{i \in \mathbb{N}^+} \mathcal{P}_i \rightarrow \mathbb{R}^n$  for  $f$  against  $\mathcal{P}$  is given by

$$Q_{f,\mathcal{P}}(\mathbf{x}, u, (\mathbf{v}_i)_i) := \beta(u) \cdot (Q_{\mathcal{P}}(\mathbf{x}/\beta(u)) - \mathbf{v}_h) + \mathbf{v}_h, \quad (7)$$

where

$$h := \min \{i : \beta(u) \cdot (Q_{\mathcal{P}}(\mathbf{x}/\beta(u)) - \mathbf{v}_i) + \mathbf{v}_i - \mathbf{x} \in L_u^+(f)\}, \quad (8)$$

and  $Q_{\mathcal{P}}$  is the lattice quantizer for basic cell  $\mathcal{P}$ .

It can be shown that LRSUQ indeed gives the desired error distribution.

**Proposition 4.** [36, Proposition 4] For any random input  $\mathbf{X}$ , the quantization error of LRSUQ  $Q_{f,\mathcal{P}}$  defined by  $\mathbf{Z} := Q_{f,\mathcal{P}}(\mathbf{X}, U, (\mathbf{V}_i)_i) - \mathbf{X}$ , follows the pdf  $f$ , independent of  $\mathbf{X}$ .

### C. Differential Privacy

The original FL framework [1] provides a certain level of privacy since clients do not directly transmit their private data to the server. However, a significant amount of information can still be inferred from the shared data, e.g., model parameters induced by gradient descent, by potential eavesdroppers within the FL network [48]–[50]. Thus, a privacy mechanism, such as DP, is necessary to be equipped on the system to protect the shared information.

In this section, we briefly review a notion of (central) DP [9], [10].<sup>6</sup> In DP, clients place trust in the server (or data curator) responsible for collecting and holding their individual data in a database  $X \in \mathcal{X}$ , where  $\mathcal{X}$  denotes the collection of databases. The server then introduces privacy-preserving noise to the original datasets or query results through a randomized mechanism  $\mathcal{F}$ , producing an output  $Y = \mathcal{F}(X) \in \mathcal{Y}$ , where  $\mathcal{Y}$  denotes the set of possible outputs, before sharing them with untrusted data analysts. While this model requires a higher level of trust compared to the local model, it enables the design of significantly more accurate algorithms. Two databases  $X$  and  $X'$  are considered *adjacent* if they differ in only one entry. More generally, we can define a symmetric adjacent relation  $\mathcal{R} \subseteq \mathcal{X}^2$  and say that  $X$  and  $X'$  are *adjacent* databases if  $(X, X') \in \mathcal{R}$ . Here, we review the definition of  $(\epsilon, \delta)$ -differentially private, initially introduced by Dwork et al. [9].

**Definition 5**  $((\epsilon, \delta)$ -differential privacy [9]). A randomized mechanism  $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$  with the associated conditional distribution  $P_{Y|X}$  of  $Y = \mathcal{F}(X)$  is  $(\epsilon, \delta)$ -differentially private ( $(\epsilon, \delta)$ -DP) if for all  $\mathcal{S} \subseteq \mathcal{Y}$  and for all  $(X, X') \in \mathcal{R}$ ,

$$\mathbb{P}(\mathcal{F}(X) \in \mathcal{S}) \leq e^\epsilon \mathbb{P}(\mathcal{F}(X') \in \mathcal{S}) + \delta.$$

When  $\delta = 0$ , we say that  $\mathcal{F}$  is  $\epsilon$ -differentially private ( $\epsilon$ -DP).

In this work, we use the principle of privacy amplification by subsampling [51], whereby the privacy guarantees of a differentially private mechanism are amplified by applying it to a random subsample of the dataset. The problem of privacy amplification can be stated as follows. Let  $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$  be a privacy mechanism with privacy profile  $\delta_{\mathcal{F}}$  with respect to the adjacent relation defined on  $\mathcal{X}$ , and let  $s : \mathcal{W} \rightarrow \mathcal{X}$  be a subsampling mechanism. Consider the subsampled mechanism  $\mathcal{F}^s : \mathcal{W} \rightarrow \mathcal{Y}$  given by  $\mathcal{F}^s(X) := \mathcal{F}(s(X))$ . The goal is to relate the privacy profile of  $\mathcal{F}$  and  $\mathcal{F}^s$ . For a detailed exposition, we refer to [51].

### D. Problem Setting

**Threat model:** We adopt trusted aggregator model, i.e., the server is trusted. Moreover, we assume that there are separate and independent sources of shared randomness between each client and the server to perform quantization. Clients engaged in the FL framework are assumed to be honest yet inquisitive, meaning they comply with the protocol but may attempt to deduce sensitive client information from the average updates received by the server. Privacy requirements are:

- Protecting the privacy of each client's local dataset from other clients, as the updated local gradients between rounds may inadvertently disclose sensitive information.
- Preventing privacy leaks from the final trained model upon completion of training, as it too may inadvertently reveal sensitive information. This ensures the relevance of our solution in scenarios where clients are trusted, and the final trained model may be publicly released to third parties.

We intend to construct a mechanism that able to jointly handle privacy requirements and compression demands (for lossless uplink channels with limited bandwidth) in a single local gradient update at each client within the FL framework. Since the distribution of the model parameters and/or the induced gradient is often unknown to the clients, our focus lies in creating a universal mechanism applicable to any random source. Furthermore, we are exploring privacy mechanisms that introduce noise customized to various noise distributions. While Laplace mechanism provides pure DP protection, Gaussian mechanism only provides approximate DP protection with a small failure probability [9], [10]. However, it is well-known that Gaussian

<sup>6</sup>Herein, DP refers to *central* DP.

mechanism support tractability of the privacy budget in mean estimation [52], [53], an important subroutine in FL. Such schemes can be formulated as mappings from the local gradient  $\mathbf{X}_t^k = \nabla F_k^{j_t^k}(\mathbf{W}_t^k) \in \mathbb{R}^m$  at client  $k$  at the time instance  $t$  to the estimated gradient  $\hat{\mathbf{X}}_t^k \in \mathbb{R}^m$  at the server, aimed to achieve the following desired properties:

- 1) Privacy requirement : The perturbed query function generated by the privacy mechanism must adhere to  $(\epsilon, \delta)$ -DP (or  $\epsilon$ -DP). For instance, ensuring that the mapping of the average of local gradients across clients  $\frac{1}{K} \sum_{k \in \mathcal{K}} \mathbf{X}_t^k$  to the average of estimated gradients  $\frac{1}{K} \sum_{k \in \mathcal{K}} \hat{\mathbf{X}}_t^k$  at the server satisfies  $(\epsilon, \delta)$ -DP.
- 2) Compression/communication-efficiency : The estimation  $\hat{\mathbf{X}}_{t+\tau}^k$  from client  $k$  to the server should be represented by finite bits per sample.
- 3) Universal source : the scheme should operate reliably irrespective of the distribution of  $\mathbf{X}_t^k$  and without prior knowledge of it.
- 4) Adaptable noise : The noise  $Z$  in the privacy mechanism is customizable according to the required accuracy level and privacy protection.

### III. CEPAM FOR FL

In this section, we recap CEPAM (with some modifications), which was initially proposed in [37]. CEPAM utilizes a randomized quantizer to provide privacy enhancement in FL setting. Specifically, CEPAM modifies schemes in [8], [20] by replacing universal quantization [39], [40] with LRSUQ as outlined in Section II-B. Moreover, CEPAM generalizes schemes in [20], [21], [35], as scalar quantization is a special case of vector quantization when the dimension is one. Using LRSUQ, as discussed in Section II-B and [35], [46], CEPAM is capable of addressing compression and privacy requirements simultaneously.

In particular, each client  $k \in \mathcal{K}$  computes the stochastic gradient  $\mathbf{X}_{t+\tau-1}^k = \nabla F_k^{j_{t+\tau-1}^k}(\mathbf{W}_{t+\tau-1}^k)$  after performing  $\tau - 1$  local SGD steps, then applies norm clipping to get  $\tilde{\mathbf{X}}_{t+\tau-1}^k$ , and applies LRSUQ to the clipped stochastic gradients at the end of each client side FL round (which is the iteration  $t + \tau - 1$ ). These quantized updates are then partitioned into vectors of suitable lengths as a set of messages  $\{(H_j^k, M_j^k)\}_{j \in \mathcal{N}}$  to the server. The server receives the messages from each client and adds dithers according to the shared randomness with each client. Next, the server decodes and collects the messages to recover  $\hat{\mathbf{X}}_t^k$  as the estimated update for each client. Subsequently, the server averages the estimated gradients from all clients as  $\frac{1}{K} \sum_{k \in \mathcal{K}} \hat{\mathbf{X}}_t^k$ , computes the new global model  $\mathbf{W}_{t+\tau}$  and broadcasts it to all clients for the next FL round, or output the global model if it is the last iteration  $T$ . A detailed pseudocode description of CEPAM is given in Algorithm 1 and a flow diagram of CEPAM is given in Figure 1.

Besides the local trainings that are carried out at each client, CEPAM can be understood as a three-stage processing: an initialization stage, an encoding stage at each client, and a decoding stage at the server. Before giving the details of each stage, we briefly introduce them. The initialization stage involves both the clients and the server before the start of the FL training, including agreement on the parameters for privacy, randomized quantizers, and the initial model. The encoding stage is independently performed by each client, where the clipped gradient is encoded into a set of messages. The decoding stage is carried out by the server, which decodes the set of messages from each client and aggregates them to recover an estimated gradient. We assume that all clients execute the same encoding function in every FL global epoch, ensuring consistency across all clients. Thus, we may focus on the  $k$ -th client in the following. The details of each stage are described below.

*1. Initialization:* Before the start of the training, client  $k$  and the server agree on the privacy parameters, involving the privacy budget  $\epsilon_k$  and privacy relaxation  $\delta_k$  for  $(\epsilon_k, \delta_k)$ -DP (or only  $\epsilon_k$  for  $\epsilon_k$ -DP), and privacy-preserving noise  $\mathbf{Z} \sim f$  with noise variance  $\text{Var}(f) > 0$ . They then align the parameters for LRSUQ as discussed in Section II-B, including a random seed  $s_k$  that serves as a source of common randomness, the lattice dimension  $n$ , a lattice generator matrix  $\mathbf{G}$ , a basic cell  $\mathcal{P}$  and a function  $\beta : (0, \infty) \rightarrow [0, \infty)$  satisfying  $L_u^+(f) \subseteq \beta(u)\mathcal{P}$  for  $u > 0$  for the randomized quantizer. Finally, both the client and the server initializes their respective PRNGs  $\mathfrak{P}$  and  $\tilde{\mathfrak{P}}$  with the shared random seed  $s_k$ , ensuring that the outputs of the two PRNGs are identical.

*2. Encoding at Client:* At the end of every FL communication round, the gradient  $\mathbf{X}_{t+\tau-1}^k \in \mathbb{R}^m$  is ready for transmission. Client  $k$  executes the encoding function ENCODE to encode the gradient into finite-bit representations by utilizing LRSUQ and entropy coding. A detailed pseudocode description of the encoding function is given in Algorithm 2. In the following we detail the quantization process:

- **Norm clipping.** Client  $k$  prepares the gradient  $\mathbf{X}_{t+\tau-1}^k \in \mathbb{R}^m$ . First, client  $k$  clips the gradient by the clipping threshold  $\gamma \in (0, M]$  to  $\tilde{\mathbf{X}}_{t+\tau-1}^k \in \mathcal{B}(0, \gamma) \subseteq \mathbb{R}^m$  so that the 2-norm of clipped gradient  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  is at most  $\gamma$ .
- **Vector partitioning.** Partition the clipped gradient  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  into  $N := \lceil \frac{m}{n} \rceil$  sub-vectors, each  $n$ -dimensional.<sup>7</sup> Denote the collection of sub-vectors by  $\{\tilde{\mathbf{X}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}} \subseteq \mathbb{R}^n$ , where  $\mathcal{N} := \{1, 2, \dots, N\}$ .
- **LRSUQ.** Let  $\mathbf{Z} \sim f$  denote the intended privacy-preserving noise random vector with mean  $\mu := \mathbb{E}[\mathbf{Z}] = \mathbf{0}$  and variance  $\text{Var}(f) := \text{Var}(\mathbf{Z})$ . Also, let  $g(u) := \mu(L_u^+(f))$  be the pdf of the latent variable corresponds to the pdf  $f$ . Then for each  $j \in \mathcal{N}$ , generate a sequence of dither signal  $\mathbf{V}_{t+\tau-1,j,1}^k, \mathbf{V}_{t+\tau-1,j,2}^k, \dots \stackrel{iid}{\sim} \text{Unif}(\mathcal{P})$  and the latent random variable

<sup>7</sup>A suitable zero-padding is required if  $m$  is not a integer multiple of  $n$ .

---

**Algorithm 1** CEPAM
 

---

1: **Inputs:** Number of total iterations  $T$ , number of local iterations  $\tau$ , number of clients  $K$ , local datasets  $\{\mathcal{D}^{(k)}\}_{k \in \mathcal{K}}$ , loss function  $\ell(\cdot, \cdot)$ , clipping threshold  $\gamma > 0$   
 2: **Output:** Global optimized model  $\mathbf{W}_T$   
 3: **Initialization:** Client  $k$  and the server agree on privacy budget  $\epsilon > 0$  and privacy relaxation  $\delta$  for  $(\epsilon, \delta)$ -DP (or privacy budget  $\epsilon > 0$  for  $\epsilon$ -DP), shared seed  $s_k$ , lattice dimension  $n$ , generator matrix  $\mathbf{G}$ , basic cell  $\mathcal{P}$  of  $\mathbf{G}\mathbb{Z}^n$ , privacy-preserving noise  $\mathbf{Z} \sim f$  with noise variance  $\text{Var}(f) > 0$ , latent variable  $U \sim g(u) := \mu(L_u^+(f))$ , function  $\beta : (0, \infty) \rightarrow [0, \infty)$  satisfying  $L_u^+(f) \subseteq \beta(u)\mathcal{P}$  for  $u > 0$ , initial model parameter vector  $\mathbf{W}_0$   
 4: **Protocol at client  $k$ :**  
 5: **for**  $t + 1 \notin \mathcal{T}_T$  **do**  
 6:   Receive  $\mathbf{W}_t$  from the server or use  $\mathbf{W}_0$  if  $t = 0$   
 7:   Set  $\mathbf{W}_t^k \leftarrow \mathbf{W}_t$   
 8:   **for**  $t' = 1$  **to**  $\tau - 1$  **do**  
 9:     Compute  $\mathbf{W}_{t+t'}^k \leftarrow \mathbf{W}_{t+t'-1}^k - \eta_{t+t'-1} \nabla F_k^{j_{t+t'-1}^k}(\mathbf{W}_{t+t'-1}^k)$   
 10:    **end for**  
 11:    Compute  $\mathbf{X}_{t+\tau-1}^k \leftarrow \nabla F_k^{j_{t+\tau}^k}(\mathbf{W}_{t+\tau-1}^k)$   
 12:    Compute  $\tilde{\mathbf{X}}_{t+\tau-1}^k \leftarrow \mathbf{X}_{t+\tau-1}^k / \max\{1, \|\mathbf{X}_{t+\tau-1}^k\|_2/\gamma\}$  ▷ Perform norm clipping  
 13:    Run subroutine ENCODE( $\mathbf{X}_{t+\tau-1}^k, s_k, \mathbf{G}, \mathcal{P}, f, \beta(u), N$ ) ▷ See Algorithm 2  
 14:    Send  $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}$  to server, using  $N \cdot (H(\text{Geom}(p(U) | U) + H(\lceil \log |\mathcal{M}(U)| \rceil | U))$  bits  
 15: **end for**  
 16: **Protocol at the server:**  
 17: **for**  $t + \tau \in \mathcal{T}_T$  **do**  
 18:   Receive  $\{H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  from clients  
 19:   **for**  $k \in \mathcal{K}$  **do**  
 20:     Run subroutine DECODE( $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}, s_k, \mathbf{G}, \mathcal{P}, f, \beta(u), N$ ) ▷ See Algorithm 3  
 21:   **end for**  
 22:   Compute  $\hat{\mathbf{W}}_{t+\tau} \leftarrow \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \tilde{\mathbf{X}}_{t+\tau-1}^k$   
 23:   Set  $\mathbf{W}_{t+\tau} \leftarrow \hat{\mathbf{W}}_{t+\tau}$  and broadcast  $\mathbf{W}_{t+\tau}$  to all clients, or output  $\mathbf{W}_T$  if  $t + \tau = T$   
 24: **end for**

---

$U_{t+\tau-1,j}^k \sim g$  according to the PRNG  $\mathfrak{P}^k$  and compute the message  $\mathbf{M}_{t+\tau-1,j}^k(i) := Q_{\mathcal{P}}(\tilde{\mathbf{X}}_{t+\tau-1,j}^k / \beta(U_{t+\tau-1,j}^k) - \mathbf{V}_{t+\tau-1,j,i}^k) \in \mathbf{G}\mathbb{Z}^n$ , where  $\beta(\cdot)$  is the deterministic scaling function defined in Definition 3. Until iteration  $i$  that satisfies  $\beta(U_{t+\tau-1,j}^k)(\mathbf{M}_{t+\tau-1,j}^k(i) + \mathbf{V}_{t+\tau-1,j,i}^k) - \tilde{\mathbf{X}}_{t+\tau-1,j}^k \in L_{U_{t+\tau-1,j}^k}^+(f)$ , take  $H_{t+\tau-1,j}^k = i$  and  $\mathbf{M}_{t+\tau-1,j}^k = \mathbf{M}_{t+\tau-1,j}^k(i)$ . Finally, transmit  $(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)$  as a binary sequence with entropy coding.

3. *Decoding at Server:* The server receives the set binary messages  $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}$  from client  $k$ . First, the server generates the same realizations of dither signals  $\mathbf{V}_{t+\tau-1,j, H_{t+\tau-1,j}^k}^k$  and latent random variables  $U_{t+\tau-1,j}^k$  according to the random seed  $s_k$  shared with client  $k$ . More specifically, for the  $j$ -th message from client  $k$ , the server uses the PRNG  $\mathfrak{P}^k$  to generate the sequence  $\mathbf{V}_{t+\tau-1,j,1}^k, \mathbf{V}_{t+\tau-1,j,2}^k, \dots, \mathbf{V}_{t+\tau-1,j, H_{t+\tau-1,j}^k}^k \stackrel{iid}{\sim} \text{Unif}(\mathcal{P})$  and  $U_{t+\tau-1,j}^k \sim g$ . Subsequently, the decoder computes the estimated sub-vector  $\mathbf{Y}_{t+\tau-1,j}^k = \beta(U_{t+\tau-1,j}^k)(\mathbf{M}_{t+\tau-1,j}^k + \mathbf{V}_{t+\tau-1,j, H_{t+\tau-1,j}^k}^k)$ . Subsequently, the server collects the sub-vectors  $\{\mathbf{Y}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  and recovers the estimated gradient  $\hat{\mathbf{X}}_{t+\tau-1}^k$ . A pseudocode description of the decoding function is given in Algorithm 3. Finally, the server computes the new global parameter vector according to:

$$\mathbf{W}_{t+\tau} \leftarrow \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \hat{\mathbf{X}}_{t+\tau-1}^k, \quad (9)$$

and broadcasts it to the clients.

In the following, we detail the privacy guarantee that is jointly ensured by the encoding-decoding steps. The privacy enhancement is ensured by the execution of LRSUQ cooperatively between the clients and the server. Through the decoding process, the estimated sub-vectors  $\{\mathbf{Y}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  are guaranteed to be noisy estimates of the clipped gradients  $\{\tilde{\mathbf{X}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$ . In other words, the estimated gradients recovered at the server are noisy versions of the clipped gradients transmitted by the clients, thereby implementing a privacy mechanism. Specifically, in Section IV-A, we prove that the global average of the estimated gradients satisfies  $(\epsilon, \delta)$ -DP requirement when using the Gaussian mechanism, and satisfies  $\epsilon$ -DP requirement when using the Laplace mechanism, respectively, by specifying appropriate pdfs  $f$  and  $g$ .

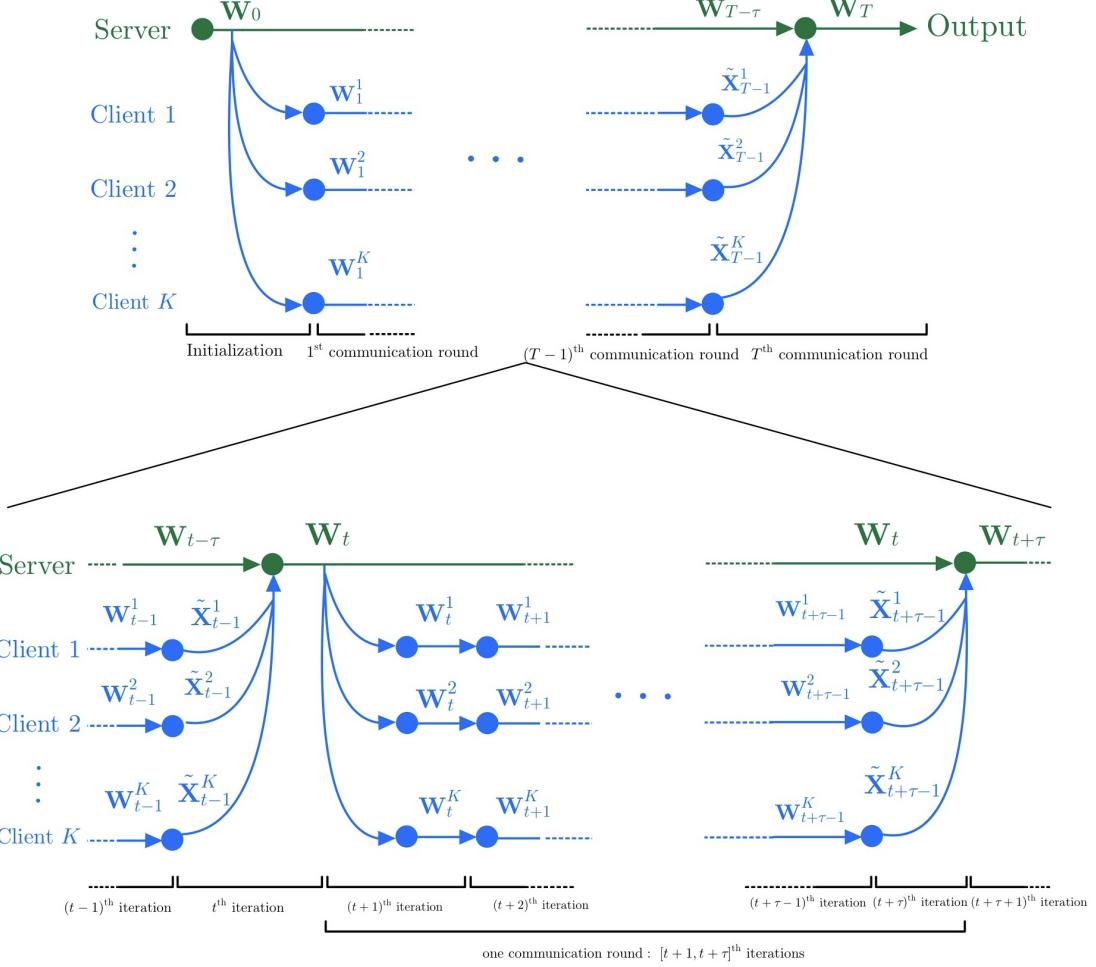


Fig. 1: Schematic of CEPAM in the FL framework. The upper part demonstrates initialization and aggregation of model updates between server and clients across global communication rounds. The lower part details one global communication round, illustrating the sequence of local updates over local iterations at clients and server-side aggregation at the end of one FL communication round. At the beginning of each communication round, client  $k$  receives the global parameter vector  $\mathbf{W}_t$  from the server and set  $\mathbf{W}_t^k = \mathbf{W}_t$ . Next, client  $k$  performs  $\tau - 1$  steps of local SGD, clips the gradient  $\nabla F_k^{j_{t+\tau}}(\mathbf{W}_{t+\tau-1}^k)$ , and encodes  $\tilde{\mathbf{X}}_{t+\tau-1}^k$ . The server then aggregates  $K$  estimated local gradients  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  and perform one step of global SGD according to  $\mathbf{W}_{t+\tau} = \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \tilde{\mathbf{X}}_{t+\tau-1}^k$ .

#### IV. PERFORMANCE ANALYSIS

In this section, we theoretically study the performance of CEPAM. We characterize the privacy guarantees and compression capabilities of CEPAM in Section IV-A and Section IV-B, respectively. Then we explore the distortion bound and convergence analysis in Section IV-C and Section IV-D, respectively.

We begin by stating a useful lemma for the subsequent analysis.

**Lemma 6.** *The LRSUQ quantization errors  $\{\tilde{\mathbf{Z}}_{t+\tau-1,j}^k := \mathbf{Y}_{t+\tau-1,j}^k - \tilde{\mathbf{X}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  are iid (over  $k$  and  $j$ ), follows the pdf  $f$ , and independent of  $\tilde{\mathbf{X}}_{t+\tau-1,j}^k$ .*

*Proof.* Since we are using LRSUQ for the encoding and decoding of the  $j$ -th sub-vector  $\tilde{\mathbf{X}}_{t+\tau-1,j}^k$  where  $j \in \mathcal{N}$ , Proposition 4 implies that, regardless of the statistical models of  $\tilde{\mathbf{X}}_{t+\tau-1,j}^k$  where  $j \in \mathcal{N}$ , the quantization errors  $\{\tilde{\mathbf{Z}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  are iid (over  $k$  and  $j$ ) and follows the pdf  $f$ .  $\square$

##### A. Privacy

By using LRSUQ, we can construct privacy mechanisms that satisfy privacy requirements by customizing  $(f, g)$ . Specifically, we present a pair of  $(f, g)$  that constructs Gaussian mechanism in Theorem 7, and a pair of  $(f, g)$  that constructs Laplace mechanism in Theorem 8, along with their associated privacy guarantees. For simplicity, we assume that  $p_k = 1/K$ . The proofs for Theorem 7 and Theorem 8 can be found in [37].

---

**Algorithm 2** ENCODE( $\tilde{\mathbf{X}}_{t+\tau-1}^k, s_k, \mathbf{G}, \mathcal{P}, f, \beta(u), N$ )

---

1: **Inputs:** Clipped gradient vector  $\tilde{\mathbf{X}}_{t+\tau-1}^k$ , random seed  $s_k$ , generator matrix  $\mathbf{G}$ , basic cell  $\mathcal{P}$  of  $\mathbf{G}\mathbb{Z}^n$ , pdf  $f$ , function  $\beta : (0, \infty) \rightarrow [0, \infty)$  satisfying  $L_u^+(f) \subseteq \beta(u)\mathcal{P}$  for  $u > 0$ , number of sub-vectors  $N$   
2: **Output:** Set of messages  $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}$   
3: Partition  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  to  $\{\tilde{\mathbf{X}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$   
4: **for**  $j = 1$  **to**  $N$  **do**

5:   Sample  $U_{t+\tau-1,j}^k \sim g$  where  $g(u) := \mu(L_u^+(f))$  by  $\mathfrak{P}^k$  ▷ Initiated by seed  $s_k$   
6:   **for**  $i = 1, 2, \dots$  **do** ▷ Perform RSUQ  
7:     Sample  $\mathbf{V}_{t+\tau-1,j,i}^k \sim \text{Unif}(\mathcal{P})$  by  $\mathfrak{P}^k$  ▷ Initiated by seed  $s_k$   
8:     Find unique  $\mathbf{M}_{t+\tau-1,j}^k \leftarrow Q_{\mathcal{P}}(\tilde{\mathbf{X}}_{t+\tau-1,j}^k / \beta(U_{t+\tau-1,j}^k) - \mathbf{V}_{t+\tau-1,j,i}^k) \in \mathbf{G}\mathbb{Z}^n$  ▷ Perform encoding  
9:     Check if  $\beta(U_{t+\tau-1,j}^k) \cdot (\mathbf{M}_{t+\tau-1,j}^k + \mathbf{V}_{t+\tau-1,j,i}^k) - \tilde{\mathbf{X}}_{t+\tau-1,j}^k \in L_{U_{t+\tau-1,j}^k}^+(f)$  ▷ Perform rejection sampling  
10:    **if** Yes **then**  
11:      $H_{t+\tau-1,j}^k \leftarrow i$ ; Return  $(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)$   
12:    **else**  
13:     Reject  $i$  and repeat Step 7-9 with  $i + 1$   
14:    **end if**  
15:   **end for**  
16: **end for**  
17: **return**  $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}$

---

**Algorithm 3** DECODE( $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}, s_k, \mathbf{G}, \mathcal{P}, f, \beta(u), N$ )

---

1: **Inputs:** Set of messages  $\{(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)\}_{j \in \mathcal{N}}$ , random seed  $s_k$ , generator matrix  $\mathbf{G}$ , basic cell  $\mathcal{P}$  of  $\mathbf{G}\mathbb{Z}^n$ , pdf  $f$ , function  $\beta : (0, \infty) \rightarrow [0, \infty)$  satisfying  $L_u^+(f) \subseteq \beta(u)\mathcal{P}$  for  $u > 0$ , number of sub-vectors  $N$   
2: **Output:** Estimated gradient  $\hat{\mathbf{X}}_{t+\tau-1}^k$   
3: Use  $\tilde{\mathfrak{P}}^k$  to sample  $U_{t+\tau-1,j}^k \sim g$  where  $g(u) := \mu(L_u^+(f))$  ▷ Initiated by seed  $s_k$   
4: **for**  $j \in \mathcal{N}$  **do**  
5:   Receive  $H_{t+\tau-1,j}^k$  and  $\mathbf{M}_{t+\tau-1,j}^k$   
6:   Use  $\tilde{\mathfrak{P}}^k$  to sample  $\mathbf{V}_{t+\tau-1,j,1}, \dots, \mathbf{V}_{t+\tau-1,j, H_{t+\tau-1,j}^k} \stackrel{iid}{\sim} \text{Unif}(\mathcal{P})$  ▷ Initiated by seed  $s_k$   
7:   Compute  $\mathbf{Y}_{t+\tau-1,j}^k \leftarrow \beta(U_{t+\tau-1,j}^k)(\mathbf{M}_{t+\tau-1,j}^k + \mathbf{V}_{H_{t+\tau-1,j}^k}^k)$  ▷ Perform decoding, inducing privacy protection  
8: **end for**  
9: Collect  $\{\mathbf{Y}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  into  $\mathbf{Y}_{t+\tau-1}^k$  and set  $\hat{\mathbf{X}}_{t+\tau-1}^k \leftarrow \mathbf{Y}_{t+\tau-1}^k$   
10: **return**  $\hat{\mathbf{X}}_{t+\tau-1}^k$

---

1) *Gaussian mechanism:* If  $g$  follows a chi-squared distribution with  $n + 2$  degrees of freedom, then  $f$  follows a Gaussian distribution. This guarantees that the global average of estimated model updates satisfies  $(\epsilon, \delta)$ -DP.

**Theorem 7.** Let  $\tau' = \tau - 1$ . Set  $\tilde{U}_{t+\tau',j}^k \sim g = \chi_{n+2}^2$  and  $\tilde{Z}_{t+\tau',j}^k | \{\tilde{U}_{t+\tau',j}^k = u\} \sim \text{Unif}(\sigma\sqrt{u}B^n)$  for every  $k, j$ , the resulting mechanism is Gaussian. The average  $\frac{1}{K} \sum_{k \in \mathcal{K}} \hat{\mathbf{X}}_{t+\tau'}^k$  is a noisy estimate of the average of the clipped gradients with

$$\frac{1}{K} \sum_{k \in \mathcal{K}} \hat{\mathbf{X}}_{t+\tau'}^k = \frac{1}{K} \sum_{k \in \mathcal{K}} \tilde{\mathbf{X}}_{t+\tau'}^k + \mathcal{N} \left( \mathbf{0}, \frac{\sigma^2}{K} \mathbf{I}_m \right),$$

where  $\mathbf{I}_m$  is the  $m \times m$  identity matrix. Hence, for every  $\tilde{\epsilon} > 0$ ,  $\frac{1}{K} \sum_{k \in \mathcal{K}} \hat{\mathbf{X}}_{t+\tau'}^k$  satisfies  $(\epsilon, \delta)$ -DP for  $\epsilon = \log(1 + p(e^{\tilde{\epsilon}} - 1))$  where  $p = 1 - \left(1 - \frac{1}{|\mathcal{D}^{(k)}|}\right)^{\tau'}$  and

$$\delta = \sum_{j=1}^{\tau'} \binom{\tau'}{j} \left( \frac{1}{|\mathcal{D}^{(k)}|} \right)^j \left(1 - \frac{1}{|\mathcal{D}^{(k)}|}\right)^{\tau'-j} \frac{(e^{\tilde{\epsilon}} - 1)}{e^{\tilde{\epsilon}/j} - 1} \times \left( \Phi \left( \frac{\tau' \gamma}{\sqrt{K} \sigma} - \frac{\sqrt{K} \tilde{\epsilon} \sigma}{2j \tau' \gamma} \right) - e^{\tilde{\epsilon}/j} \Phi \left( -\frac{\tau' \gamma}{\sqrt{K} \sigma} - \frac{\sqrt{K} \tilde{\epsilon} \sigma}{2j \tau' \gamma} \right) \right). \quad (10)$$

2) *Laplace mechanism:* If  $g$  follows a Gamma distribution  $\text{Gamma}(2, 1)$ , then  $f$  follows a Laplace distribution. This guarantees that estimated model updates satisfy  $\epsilon$ -DP.

**Theorem 8.** Let  $\tau' = \tau - 1$ . Set  $\tilde{U}_{t+\tau',j}^k \sim g = \text{Gamma}(2, 1)$  and  $\tilde{Z}_{t+\tau',j}^k | \{\tilde{U}_{t+\tau',j}^k = u\} \sim \text{Unif}((-bu, bu))$  for every  $k$  and  $j$ , the resulting mechanism is Laplace. The estimator  $\hat{\mathbf{X}}_{t+\tau'}^k$  is a noisy estimate of the clipped gradients  $\tilde{\mathbf{X}}_{t+\tau'}^k$  such that

$$\hat{\mathbf{X}}_{t+\tau'}^k = \tilde{\mathbf{X}}_{t+\tau'}^k + \text{Lap}(\mathbf{0}, b\mathbf{I}_m), \quad (11)$$

where  $\mathbf{I}_m$  is the  $m \times m$  identity matrix. Hence, for every  $\hat{\mathbf{X}}_{t+\tau}^k$  satisfies  $(\epsilon, 0)$ -DP in one round against clients for  $\epsilon = \log(1 + p(e^{\tilde{\epsilon}} - 1))$  where  $p = 1 - \left(1 - \frac{1}{|\mathcal{D}^{(k)}|}\right)^{\tau'}$ , and  $\delta = 0$  provided that  $\tilde{\epsilon} \geq 2\tau'\gamma/b$ .

### B. Compression

Every client in CEPAM is required to transmit the set of message pairs  $(H_{t+\tau-1,j}^k, \mathbf{M}_{t+\tau-1,j}^k)_{j \in \mathcal{N}}$  per communication round. If we know that  $\tilde{\mathbf{X}}_{t+\tau-1,j}^k \in \mathcal{X}$ , then we can compress  $H_{t+\tau-1,j}^k$  using the optimal prefix-free code [54], [55] for  $\text{Geom}(p(u))|\{U_{t+\tau-1,j}^k = u\}$ , where  $p(u) := \mu(L_u^+(f))/\mu(\beta(u)\mathcal{P})$ , and compress  $\mathbf{M}_{t+\tau-1,j}^k|\{U_{t+\tau-1,j}^k = u\} \in \mathcal{M} := (\mathcal{X} + L_u^+(f) - \beta(u)\mathcal{P}) \cap \mathbf{GZ}^n$  using  $H(\lceil \log |\mathcal{M}(U)| \rceil | U)$  bits. Therefore, the total communication cost per communication round per client is at most  $N \cdot (H(\text{Geom}(p(U) | U) + H(\lceil \log |\mathcal{M}(U)| \rceil | U))$  bits.

### C. Distortion Bounds

We consider a clipped version of FedAvg. This is achieved by modifying Equation 4 as follows:

$$\tilde{\mathbf{W}}_{t+\tau} \leftarrow \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \tilde{\mathbf{X}}_{t+\tau-1}^k. \quad (12)$$

Representing the clipped gradient  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  of client  $k$  using a finite number of bits inherently introduces distortion, as for the recovered vector  $\hat{\mathbf{X}}_{t+\tau-1}^k = \tilde{\mathbf{X}}_{t+\tau-1}^k + \mathbf{Z}_{t+\tau-1}^k$ , where  $\mathbf{Z}_{t+\tau-1}^k := (\tilde{\mathbf{Z}}_{t+\tau-1,1}^k, \dots, \tilde{\mathbf{Z}}_{t+\tau-1,N}^k)$ . We can then characterize the squared norm of the FL quantization error. The proof is given in Appendix A.

**Proposition 9.** *The FL quantization error  $\mathbf{Z}_{t+\tau-1}^k$  has zero mean and satisfies*

$$\mathbb{E}[\|\mathbf{Z}_{t+\tau-1}^k\|^2 | \tilde{\mathbf{x}}_{t+\tau-1}^k] = N\text{Var}(f). \quad (13)$$

At the end of each FL round, CEPAM incorporates a DP mechanism based on LRSUQ, inherently introducing distortion as mentioned above. This distortion emerges as the clipped gradient  $\tilde{\mathbf{X}}_{t+\tau-1}^k$  is mapped into its distorted counterpart  $\hat{\mathbf{X}}_{t+\tau-1}^k$ . Consequently, the global model (12) becomes

$$\hat{\mathbf{W}}_{t+\tau} := \mathbf{W}_t - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \hat{\mathbf{X}}_{t+\tau-1}^k. \quad (14)$$

Under common assumptions used in FL analysis, we can bound the error between  $\mathbf{W}_{t+\tau}$  and  $\hat{\mathbf{W}}_{t+\tau}$ . We have the following bound on the error between  $\mathbf{W}_{t+\tau}$  and  $\hat{\mathbf{W}}_{t+\tau}$ . The proof is given in Appendix B.

**Proposition 10.** *The mean-squared error  $\mathbb{E}[\|\hat{\mathbf{W}}_{t+\tau} - \mathbf{W}_{t+\tau}\|^2]$  is upper-bounded by*

$$\eta_{t+\tau-1}^2 N(\text{Var}(f) + M^2) \sum_{k \in \mathcal{K}} p_k^2. \quad (15)$$

### D. FL Convergence Analysis

Next, we study the FL convergence of CEPAM, considering common assumptions used in the FL [8], [38], [56]:

**AS1:** The  $\xi_{k,j}$ 's, where  $j = 1, \dots, n_k$ , are i.i.d. samples in  $\mathcal{D}^{(k)}$ . However, different datasets  $\mathcal{D}^{(k)}$  can be statistically heterogeneous and follow different distributions.

**AS2:** Let the sample index  $j_t^k$  be sampled from client  $k$ 's local data  $\mathcal{D}^{(k)}$  uniformly at random. The expected squared norm of stochastic gradients is bounded by some  $\theta_k^2 > 0$ , i.e.,  $\mathbb{E}[\|\nabla F_k^{j_t^k}(\mathbf{W})\|^2] \leq \theta_k^2$ , for all  $\mathbf{W} \in \mathbb{R}^m$ . It is clear that  $\max_{k \in \mathcal{K}} \theta_k \leq M$ .

**AS3:**  $F_1, \dots, F_K$  are all  $L$ -smooth, i.e., for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ ,  $F_k(\mathbf{y}) - F_k(\mathbf{x}) \leq (\mathbf{y} - \mathbf{x})^T \nabla F_k(\mathbf{x}) + \frac{L}{2} \|\mathbf{y} - \mathbf{x}\|^2$ .

**AS4:**  $F_1, \dots, F_K$  are all  $C$ -strongly convex,<sup>8</sup> i.e., for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^m$ ,  $F_k(\mathbf{y}) - F_k(\mathbf{x}) \geq (\mathbf{y} - \mathbf{x})^T \nabla F_k(\mathbf{x}) + \frac{C}{2} \|\mathbf{y} - \mathbf{x}\|^2$ .

We define the following to quantify the degree of non-iid:

$$\psi := F(\mathbf{w}^*) - \sum_{k \in \mathcal{K}} \min_{\mathbf{w} \in \mathbb{R}^m} F_k(\mathbf{w}), \quad (16)$$

where  $\mathbf{w}^*$  denotes the minimum of  $F(\cdot)$ . If the data is iid, meaning that the training data originates from the same distribution, then  $\psi$  goes to zero as the training size increases. In the case of non-i.i.d. data,  $\psi$  is positive, and its magnitude reflects the heterogeneity of the data distribution.

The convergence of CEPAM for FL learning is given as follows. The proof is given in Appendix C.

<sup>8</sup>The extension to non-convex objective functions is left for future work.

**Theorem 11.** Assume that AS1-4 hold and  $\theta_k L, C$  be defined therein. Set  $\alpha = \tau \max\{\frac{4L}{C}, 1\}$  and the learning rate  $\eta_t = \frac{\tau}{C(t+\alpha)}$  for all  $t \in \mathcal{T}$ . Then, CEPAM satisfies

$$\mathbb{E}[F(\mathbf{W}_T)] - F(\mathbf{w}^*) \leq \frac{L}{2(T+\alpha)} \max \left\{ \frac{4B\tau^2(\tau+\alpha)}{C^2\alpha(\tau-1)}, \alpha \|\mathbf{W}_0 - \mathbf{w}^*\| \right\}, \quad (17)$$

where

$$B := 6L\psi + N(\text{Var}(f) + M^2) \sum_{k \in \mathcal{K}} p_k^2 + \sum_{k \in \mathcal{K}} p_k^2 \theta_k^2 + 8(\tau-1)^2 \sum_{k \in \mathcal{K}} p_k \theta_k^2.$$

A decreasing upper bound for all  $t \geq 0$  can also be obtained. Between two consecutive global communication rounds, the server first broadcasts the model to all clients, and each client performs multiple local updates. Under Assumptions **AS1-AS4**, these local iterations constitute standard stochastic gradient steps, and standard gradient-descent analysis ensures that the expected objective value decreases across iterations, yielding convergence for all  $t \geq 0$ . However, to establish convergence of the model to the true minimizer, it suffices to analyze the process at the synchronous global communication rounds.

## V. EXPERIMENTAL EVALUATIONS

In this section, we evaluate the performance of CEPAM.<sup>9</sup> We begin by detailing our experimental setup, which includes the types of datasets, model architectures, and training configurations in Section V-A. In Section V-B, we present comprehensive experimental results to demonstrate the effectiveness of CEPAM by comparing it to several baselines that are commonly used for privacy protection and quantization in the FL framework. In addition, we investigate the accuracy-privacy trade-off phenomenon for CEPAM-Gaussian and CEPAM-Laplace. All experiments were executed on a server equipped with dual Intel Xeo nGold 6326 CPUs (48 cores and 96 threads in total), 256 GiB RAM, and two NVIDIA RTX A6000 GPUs (each with 48GB VRAM), running Ubuntu 22.04.5 LTS. The implementation was based on Python 3.13.5 and PyTorch 2.7.1+cu126, with CUDA 11.5 and NVIDIA driver version 565.57.01.

### A. Experimental Setup

1) *Datasets:* We evaluate CEPAM on the standard image classification benchmark MNIST, which consists of  $28 \times 28$  grayscale handwritten digits image divided into 60,000 training examples and 10,000 testing examples. The training examples and the testing examples are equally distributed among  $K = 30$  clients. For simplicity, we set  $p_k = 1/K$  for all  $k \in \mathcal{K}$ .

2) *Learning Architecture:* We evaluate CEPAM using the convolutional neural network (CNN), which is composed of two convolutional layers followed by two fully-connected ones, with intermediate ReLU activations and max-pooling layers. Also, the model uses a softmax output layer. There are 6422 learnable parameters for CNN.

3) *Baselines:* We compare CEPAM to the following baselines:

- FL: vanilla FL without any privacy or compression scheme.
- FL+SDQ: FL with scalar SDQ-based compression but no privacy scheme.
- FL+{Gaussian, Laplace}: FL with one-dimensional Gaussian or Laplace mechanism, but no compression scheme.
- FL+{Gaussian, Laplace}+SDQ: FL with approach that applies one-dimensional Gaussian or Laplace mechanism followed by SDQ.
- CEPAM-{Gaussian, Laplace}: CEPAM achieves privacy and compression jointly through LRSUQ to construct Gaussian or Laplace mechanisms, namely CEPAM-Gaussian or CEPAM-Laplace. In particular, we evaluate three different cases of LRSUQ that simulate Gaussian noise for dimension  $n = 1, 2, 3$ , where we use the scaled integer lattice  $\alpha\mathbb{Z}^n$  with  $\alpha = 10^{-3}$ , and the corresponding basic cells are  $\alpha(-0.5, 0.5]$ ,  $\alpha(-0.5, 0.5]^2$ , and  $\alpha(-0.5, 0.5]^3$ . We have chosen  $\alpha = 10^{-3}$  as the precision level for quantization as it aligns with the observed precision level of the model updates during training.

4) *Training Configurations:* We select the momentum SGD as the optimizer, where the momentum is set to 0.9. The local iterations  $\tau$  per global epoch is set to 15. Also, we set the initial learning rate to be 0.1.

5) *Repetition Strategy:* For each set of parameters and baseline method, the training process was repeated 10 times with varying random seeds. Indicators of performance, including average and 95% confidence interval across these runs was reported in the following sections.

### B. FL Convergence

We report the FL convergence of CEPAM-Gaussian and CEPAM-Laplace in terms of accuracy using CNN architecture and MNIST dataset.

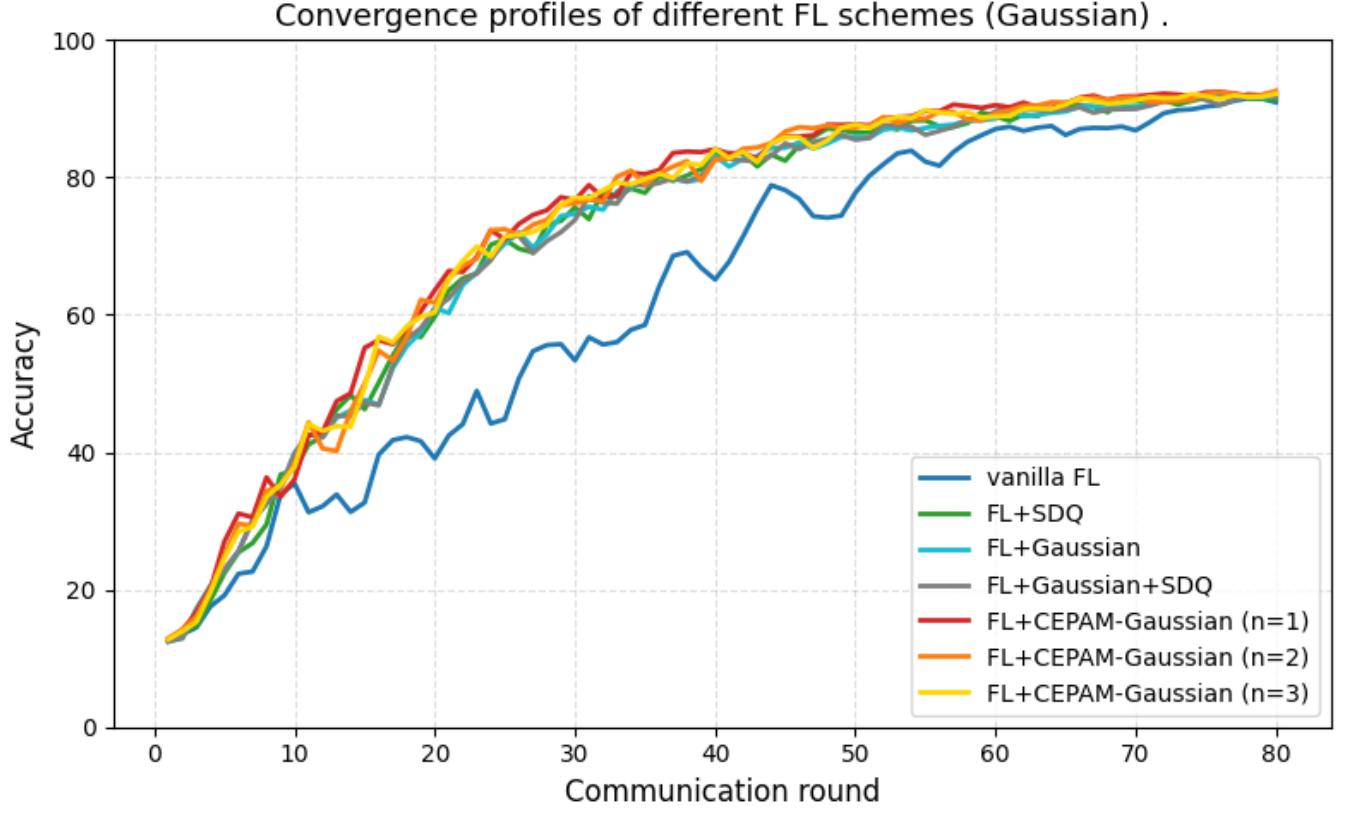


Fig. 2: Convergence profile of different FL schemes for Gaussian

1) *Gaussian Mechanism*: We set  $\sigma = 0.01$  and  $\tilde{\epsilon} = 5.9$  for the base Gaussian mechanisms in CEPAM-Gaussian for  $n = 1, 2, 3$  with variance  $0.01^2$ ,  $2 \times 0.01^2$  and  $3 \times 0.01^2$  respectively. By Theorem 7, all the clients' composite Gaussian mechanisms achieve  $(\epsilon = 1.45, \delta = 9.69 \times 10^{-3})$ -DP. We also set  $\sigma = 0.01$  for both FL+Gaussian and FL+Gaussian+SDQ.

Figure 2 shows the validation accuracy of CEPAM-Gaussian over 80 global communication rounds using the CNN model. Note that the cases of CEPAM-Gaussian ( $n = 1, 2, 3$ ) outperform and achieve higher accuracy than all the other baselines. The pronounced fluctuations observed in the figure are mainly attributed to two factors. First, the server aggregates raw stochastic gradient at each communication round, instead of aggregating model differences in [37], which inherently exhibit high variance. Second, a fixed learning rate is employed throughout the entire training phase.

Baselines	Accuracy (%)	Average SNR (dB)
FL	$93.24 \pm 0.31$	$\infty$
FL+SDQ	$93.28 \pm 0.63$	84.22
FL+Gaussian	$93.46 \pm 0.70$	24.52
FL+Gaussian+SDQ	$93.18 \pm 0.59$	24.65
CEPAM-Gaussian ( $n = 1$ )	$94.11 \pm 0.49$	8.15
CEPAM-Gaussian ( $n = 2$ )	$94.06 \pm 0.45$	15.26
CEPAM-Gaussian ( $n = 3$ )	$93.86 \pm 0.36$	19.76

TABLE I: Test Accuracy for Gaussian on MNIST

In Table I, we report the test accuracy of CEPAM-Gaussian with their 95% confidence intervals and the average SNR ratio (in dB). CEPAM-Gaussian in different dimensions all demonstrate better accuracy, achieving an improvement of 0.4-0.9% in accuracy compared to the other baselines. Furthermore, CEPAM-Gaussian of different dimension have approximately the same accuracy. This is because the added noise per dimension is consistent, i.e.,  $\text{Var}(f)/n = \sigma^2$ . However, higher-dimension LRSUQ has a better compression ratio than scalar counterparts [20], [21], due to its nature as a vector quantizer.

2) *Laplace Mechanism*: We set  $b = 0.01$  and  $\tilde{\epsilon} = 3000$  for the base Laplace mechanism in CEPAM-Laplace with a variance  $2 \times 0.01^2$ , and the composite Laplace mechanism achieves  $(\epsilon = 2995)$ -DP. We also set  $b = 0.01$  for both FL+Laplace and FL+Laplace+SDQ. Note that a higher privacy budget is required to ensure  $\epsilon$ -DP for CEPAM-Laplace. However, we might achieve a lower privacy budget by relaxing  $\delta$ . The trade-off of this relaxation is left for future study.

<sup>9</sup>The source code used in our numerical evaluations is available at <https://github.com/shiushiu4863/CEPAM-Gradient>.

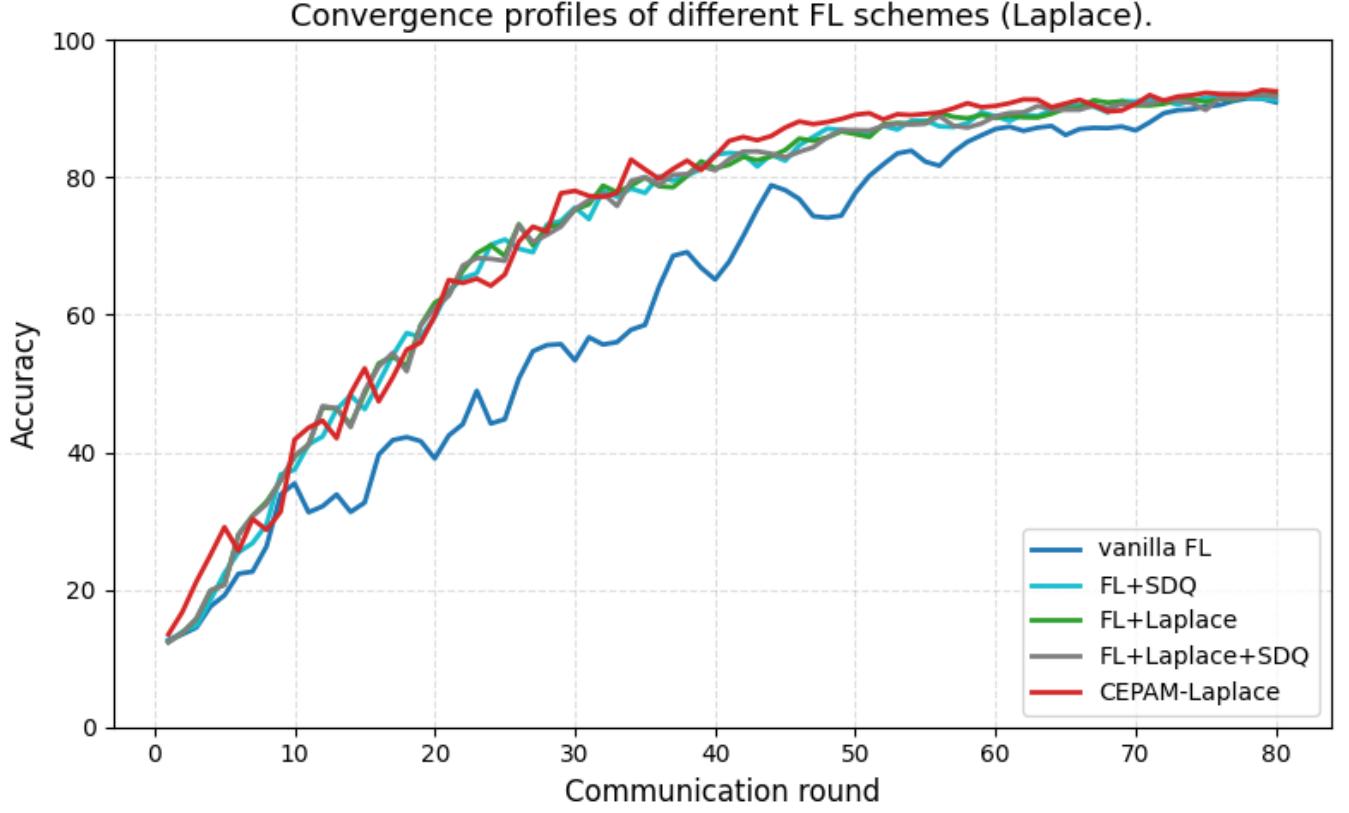


Fig. 3: Convergence profile of different FL schemes for Laplace

Figure 3 shows the validation accuracy of CEPAM-Laplace over global epochs using the CNN model. Similar to CEPAM-Gaussian, we observe that CEPAM-Laplace outperforms other baselines.

Baselines	Accuracy (%)	Average SNR (dB)
FL	93.24 $\pm$ 0.31	$\infty$
FL+SDQ	93.28 $\pm$ 0.63	84.22
FL+Laplace	93.43 $\pm$ 0.54	20.96
FL+Laplace+SDQ	93.25 $\pm$ 0.59	21.21
CEPAM-Laplace	94.32 $\pm$ 0.40	-1.582

TABLE II: Test Accuracy for Laplace on MNIST

In Table II, we also report the performance of test accuracy of CEPAM-Laplace with their 95% confidence intervals and the average SNR ratio (in dB). We observe that CEPAM-Laplace achieves an improvement of about 0.8-1.1% in accuracy compared to the other baselines.

It is worth noting that adding a minor level of distortion during training deep models can potentially enhance the performance of the converged model [57]. This finding corroborates in our experimental results, as both CEPAM-Gaussian and CEPAM-Laplace achieves slightly better accuracy than vanilla FL.

### C. Privacy-Accuracy Trade-off

1) *Gaussian Mechanism:* We evaluate the privacy-accuracy trade-off with CEPAM-Gaussian. In the following, the parameters are computed according to Theorem 7. We set  $\delta = 0.015$ , we conducted the experiments by varying privacy budget  $\epsilon$  from 0.5 to 5, with step size 0.5. The corresponding noise levels  $\sigma$  are  $\{0.13881, 0.08567, 0.06058, 0.04447, 0.03274, 0.02362, 0.01623, 0.01011, 0.00494, 0.00052\}$  respectively. For each set of parameters, the experiments are simulated 10 times with different random seeds, and the results are averaged.

Figure 4 illustrates the trade-off between learning performance, measured in terms of test accuracy, and the privacy budget between CEPAM-Gaussian and the simple Gaussian-mechanism-then-quantize approach (Gaussian+SDQ).

Overall, CEPAM-Gaussian has a better performance than Gaussian+SDQ. The figure demonstrates that as more privacy budget is allocated, higher test accuracy can be achieved with CEPAM-Gaussian. However, there is a point of diminishing

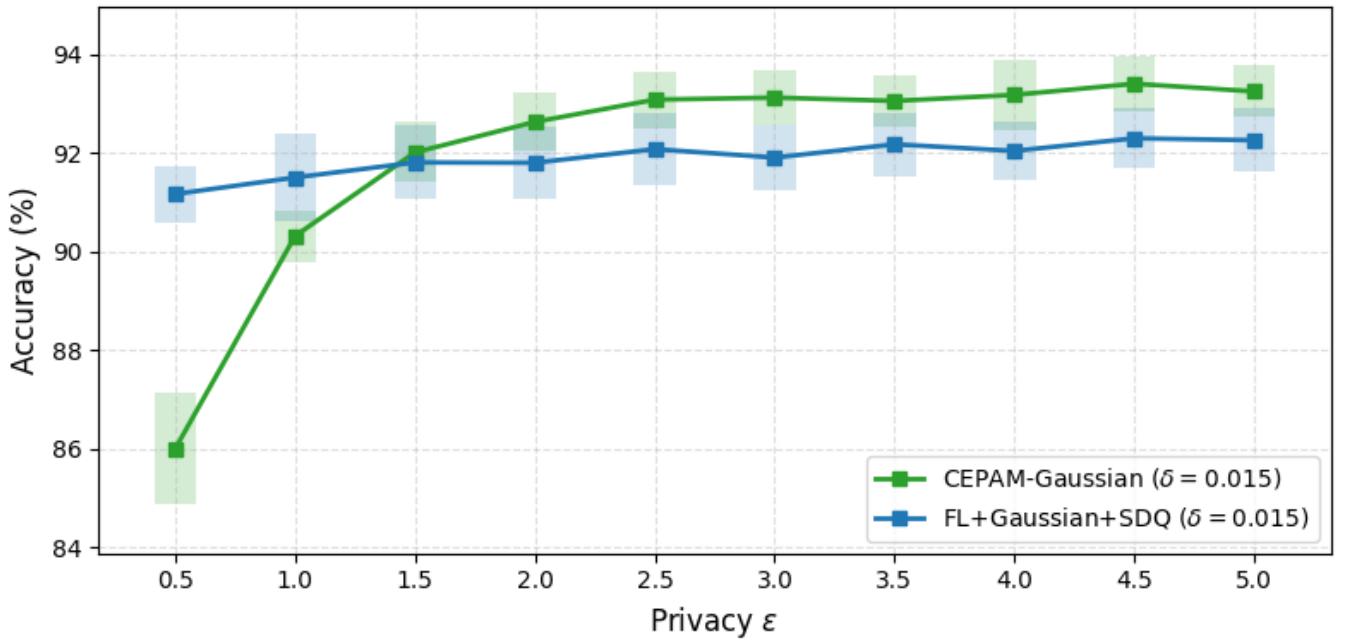


Fig. 4: Accuracy and Privacy Trade-off of Gaussian

returns; once the privacy budget reaches a threshold ( $\epsilon \approx 2.5$ ), the increase in test accuracy by further increasing the privacy budget becomes limited.

2) *Laplace Mechanism:* Similarly, we evaluate the privacy-accuracy trade-off with CEPAM-Laplace. In the following, the parameters are computed according to Theorem 8. We conducted the experiments by varying privacy budget  $\epsilon$  from 500 to 5000, with step size 500. The corresponding noise levels  $b$  are  $\{0.05944, 0.029859, 0.019937, 0.014965, 0.011977, 0.009984, 0.00856, 0.007491, 0.00666, 0.005994\}$  respectively. Again, for each set of parameters, the experiments are simulated 10 times with different random seeds, and the results are averaged.

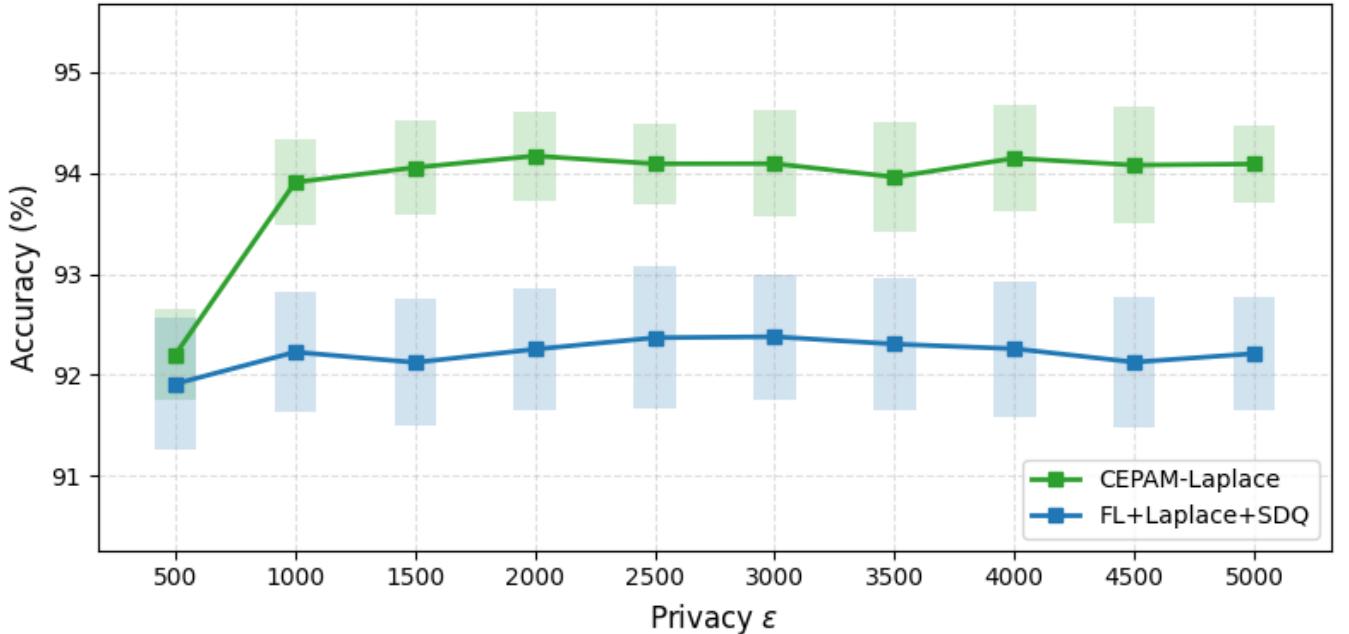


Fig. 5: Accuracy and Privacy Trade-off of Laplace

Figure 5 illustrates the trade-off between learning performance, measured in terms of test accuracy, and the privacy budget between CEPAM-Laplace and the simple Laplace-mechanism-then-quantize approach (Laplace+SDQ).

Overall, CEPAM-Laplace outperforms Gaussian+SDQ. A similar diminishing return phenomenon can be observed in this case, as the privacy budget reaches a threshold ( $\epsilon \approx 1500$ ), the increase in test accuracy by further increasing the privacy budget becomes limited.

It is worth noting that the test accuracy slightly decreased when we further increase the privacy budget ( $\epsilon \approx 4000$ ), i.e., lowering the noise level  $b$ . This finding again coincides with the finding that adding a minor level of distortion can potentially enhance performance.

## VI. CONCLUSION AND DISCUSSION

In this paper, we further investigated CEPAM, which is designed to achieve communication efficiency and privacy protection simultaneously in the FL framework. We proved the FL convergence bound for CEPAM. We also conducted various experiments on a modified version of CEPAM, including convergence profiles and accuracy-privacy trade-off behavior. The experimental results confirmed our theoretical findings and demonstrated that both CEPAM-Gaussian and CEPAM-Laplace lead to an enhancement in test accuracy compared to several other commonly used baselines in the FL framework.

For potential future work, several interesting directions emerge. One could be to conduct a convergence analysis for the nonconvex case and extend the privacy analysis to other subsampling mechanisms. Another promising avenue would be to adapt CEPAM for personalized federated learning and rigorously evaluate its performance and efficacy.

## ACKNOWLEDGMENT

The authors thank Ms. Youqi Wu for developing the original framework for the simulations in Section V.

## APPENDIX A PROOF OF PROPOSITION 9

Fix  $\tilde{\mathbf{X}}_{t+\tau-1,j}^k = \tilde{\mathbf{x}}_{t+\tau-1,j}^k$ . By Lemma 6, the LRSUQ quantization errors  $\{\tilde{\mathbf{Z}}_{t+\tau-1,j}^k := \mathbf{Y}_{t+\tau-1,j}^k - \tilde{\mathbf{x}}_{t+\tau-1,j}^k\}_{j \in \mathcal{N}}$  are iid (over  $k$  and  $j$ ) and follows the pdf  $f$ . Hence,

$$\begin{aligned}\mathbb{E}[\|\mathbf{Z}_{t+\tau-1}^k\|^2 | \tilde{\mathbf{x}}_{t+\tau-1}^k] &= \sum_{j \in \mathcal{N}} \mathbb{E}[\|\tilde{\mathbf{Z}}_{t+\tau-1,j}^k\|^2] \\ &\stackrel{(a)}{=} \sum_{j \in \mathcal{N}} \text{Var}(f) \\ &= N\text{Var}(f),\end{aligned}$$

where (a) is because the mean of  $\tilde{\mathbf{Z}}_{t+\tau-1,j}^k \sim f$  is  $\mathbf{0}$ . This completes the proof.

## APPENDIX B PROOF OF PROPOSITION 10

Let  $\{\mathbf{W}_{t+\tau,j}\}_{j \in \mathcal{N}}$  and  $\{\hat{\mathbf{W}}_{t+\tau,j}\}_{j \in \mathcal{N}}$  denote the partitions of  $\mathbf{W}_{t+\tau}$  and  $\hat{\mathbf{W}}_{t+\tau}$  into  $N$  distinct  $n$ -dimensional sub-vectors, respectively. We have, for  $j \in \mathcal{N}$ ,

$$\hat{\mathbf{W}}_{t+\tau,j} = \mathbf{W}_{t,j} - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \hat{\mathbf{X}}_{t+\tau-1,j}^k,$$

and

$$\mathbf{W}_{t+\tau,j} = \mathbf{W}_{t,j} - \eta_{t+\tau-1} \sum_{k \in \mathcal{K}} p_k \mathbf{X}_{t+\tau-1,j}^k.$$

Hence,

$$\begin{aligned}\mathbb{E}[\|\hat{\mathbf{W}}_{t+\tau} - \mathbf{W}_{t+\tau}\|^2] &= \mathbb{E}\left[\left\|\eta_{t+\tau-1} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} p_k (\hat{\mathbf{X}}_{t+\tau-1,j}^k - \mathbf{X}_{t+\tau-1,j}^k)\right\|^2\right] \\ &\stackrel{(a)}{\leq} \eta_{t+\tau-1}^2 \sum_{k \in \mathcal{K}} p_k^2 \sum_{j \in \mathcal{N}} \mathbb{E}\left[\|\hat{\mathbf{X}}_{t+\tau-1,j}^k - \mathbf{X}_{t+\tau-1,j}^k\|^2\right] \\ &= \eta_{t+\tau-1}^2 \sum_{k \in \mathcal{K}} p_k^2 \sum_{j \in \mathcal{N}} \mathbb{E}\left[\|(\hat{\mathbf{X}}_{t+\tau-1,j}^k - \tilde{\mathbf{X}}_{t+\tau-1,j}^k) + (\tilde{\mathbf{X}}_{t+\tau-1,j}^k - \mathbf{X}_{t+\tau-1,j}^k)\|^2\right] \\ &\leq \eta_{t+\tau-1}^2 \sum_{k \in \mathcal{K}} p_k^2 \sum_{j \in \mathcal{N}} \left(\mathbb{E}\left[\|(\hat{\mathbf{X}}_{t+\tau-1,j}^k - \tilde{\mathbf{X}}_{t+\tau-1,j}^k)\|^2\right] + \mathbb{E}\left[\|(\tilde{\mathbf{X}}_{t+\tau-1,j}^k - \mathbf{X}_{t+\tau-1,j}^k)\|^2\right]\right) \\ &\leq \eta_{t+\tau-1}^2 \sum_{k \in \mathcal{K}} p_k^2 \sum_{j \in \mathcal{N}} \left(\mathbb{E}\left[\|\tilde{\mathbf{Z}}_{t+\tau-1,j}^k\|^2\right] + M^2\right) \\ &\stackrel{(b)}{=} \eta_{t+\tau-1}^2 N \sum_{k \in \mathcal{K}} p_k^2 (\text{Var}(f) + M^2),\end{aligned}$$

where (a) is by the triangle inequality, and (b) is by the argument of Proposition 9. This completes the proof.

APPENDIX C  
PROOF OF THEOREM 11

Our proof adopts similar strategies to those in [8], [38], [56], supplemented by additional arguments to address the LRSUQ quantization error. The distinctive features of the quantization error introduced by LRSUQ, as discussed in Lemma 6, enable us to establish the convergence property of CEPAM. We include the proof for the sake of completeness.

Recall that  $\mathbf{Z}_{t-1}^k$  where  $t \in \mathcal{T}_T$  is the FL quantization error, which is independent of the clipped gradient  $\tilde{\mathbf{X}}_{t-1}^k$  by Lemma 6. Let  $\mathbf{E}_{t-1}^k := \mathbf{Z}_{t-1}^k + \mathbf{C}_{t-1}^k$  if  $t \in \mathcal{T}_T$  and  $\mathbf{E}_t^k := \mathbf{0}$  otherwise, where  $\mathbf{C}_{t-1}^k$  is the clipping error, then the update of CEPAM can also be described alternatively (compared to (3) and (14)) as follows:

$$\mathbf{W}_{t+1}^k = \begin{cases} \mathbf{W}_t^k - \eta_t \nabla F_k^{j_t^k}(\mathbf{W}_t^k) + \mathbf{E}_t^k & \text{if } t+1 \notin \mathcal{T}_T, \\ \sum_{k' \in \mathcal{K}} p_{k'} \left( \mathbf{W}_{t+1-\tau}^{k'} - \eta_t \nabla F_{k'}^{j_t^{k'}}(\mathbf{W}_t^{k'}) + \eta_t \mathbf{E}_t^{k'} \right) & \text{if } t+1 \in \mathcal{T}_T. \end{cases} \quad (18)$$

By applying the strategy outlined in [38] and adapting it to Assumption 1, i.e., heterogeneous dataset as discussed in [56], we define a virtual sequence  $(\mathbf{W}'_t)_{t \in \{0\} \cup [T]}$ , where  $[T] := \{1, 2, \dots, T\}$ , from the FL model weights  $(\mathbf{W}_t^k)_{t \in \{0\} \cup [T]}$  as follows:

$$\mathbf{W}'_t = \sum_{k \in \mathcal{K}} p_k \mathbf{W}_t^k, \quad (19)$$

which coincides with  $\mathbf{W}_t^k$  when  $t \in \mathcal{T}_T$ . This sequence can be demonstrated to behave similarly to mini-batch SGD with a batch size of  $\tau$  and being bounded within a bounded distance of  $(\mathbf{W}_t^k)_{t \in \{0\} \cup [T]}$ , by appropriately configuring the step size  $\eta_t$  [38]. For convenience, define the *averaged full gradients* and the *averaged perturbed stochastic gradients* as

$$\bar{\mathbf{J}}_t := \sum_{k \in \mathcal{K}} p_k \nabla F_k(\mathbf{W}_t^k), \quad (20)$$

$$\mathbf{J}_t := \sum_{k \in \mathcal{K}} p_k \left( \nabla F_k^{j_t^k}(\mathbf{W}_t^k) - \mathbf{E}_t^k \right), \quad (21)$$

respectively. Note that since the FL quantization error  $\mathbf{Z}_{t+\tau-1}$  has zero mean by Proposition 9 and the sample indices  $j_t^k$  are independent and uniformly distributed, it follows that  $\mathbb{E}[\mathbf{J}_t] = \bar{\mathbf{J}}_t$ . Moreover, the virtual sequence satisfies  $\mathbf{W}'_{t+1} = \mathbf{W}'_t - \eta_t \mathbf{J}_t$  if  $t+1 \notin \mathcal{T}_T$  and  $\mathbf{W}'_{t+1} = \mathbf{W}'_{t+1-\tau} - \eta_t \mathbf{J}_t$  if  $t+1 \in \mathcal{T}_T$  (see Figure 1).

Therefore, the resulting model is equivalent to the model discussed in [56, Appendix A] or [8, Appendix C]. This implies that we can apply the following result (one step SGD for heterogeneity dataset) from [56, Lemma 1] for  $t+1 \notin \mathcal{T}_T$ :

$$\mathbb{E}[\|\mathbf{W}'_{t+1} - \mathbf{w}^*\|^2] \leq (1 - \eta_t C) \mathbb{E}[\|\mathbf{W}'_t - \mathbf{w}^*\|^2] + 6L\eta_t^2\psi + \eta_t^2 \mathbb{E}[\|\mathbf{J}_t - \bar{\mathbf{J}}_t\|^2] + 2\mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|\mathbf{W}'_t - \mathbf{W}_t^k\|^2 \right], \quad (22)$$

provided that  $\eta_t \leq \frac{1}{4L}$  and Assumptions 3-4 hold, and for  $t+1 \in \mathcal{T}_T$ :

$$\mathbb{E}[\|\mathbf{W}'_{t+1} - \mathbf{w}^*\|^2] \leq (1 - \eta_t C) \mathbb{E}[\|\mathbf{W}'_{t+1-\tau} - \mathbf{w}^*\|^2] + 6L\eta_t^2\psi + \eta_t^2 \mathbb{E}[\|\mathbf{J}_t - \bar{\mathbf{J}}_t\|^2] + 2\mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|\mathbf{W}'_t - \mathbf{W}_t^k\|^2 \right], \quad (23)$$

provided that  $\eta_t \leq \frac{1}{4L}$  and Assumptions 3-4 hold.

The third term in RHS of (22) and of (23) can be upper-bounded by using the following lemma, where the proof is given in Appendix D:

**Lemma 12.** *When AS2 holds, we have*

$$\mathbb{E}[\|\bar{\mathbf{J}}_t - \mathbf{J}_t\|^2] \leq \sum_{k \in \mathcal{K}} p_k^2 (\theta_k^2 + N(\text{Var}(f) + M^2)). \quad (24)$$

Furthermore, the last term in RHS of (22) and of (23) can be upper-bounded by using the following lemma, where the proof is given in Appendix E:

**Lemma 13.** *Assume that  $\eta_t$  is non-increasing and  $\eta_t \leq 2\eta_{t+\tau}$  for all  $t \geq 0$ . When Assumption 2 holds, we have*

$$\mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|\mathbf{W}'_t - \mathbf{W}_t^k\|^2 \right] \leq 4(\tau - 1)^2 \eta_t^2 \sum_{k \in \mathcal{K}} p_k \theta_k^2. \quad (25)$$

Therefore, by defining  $\delta_t := \mathbb{E}[\|\mathbf{W}'_t - \mathbf{w}^*\|]$  and substituting (24) and (25) into (22) and (23), we have the following recursive bounds:

$$\delta_{t+1} \leq (1 - C\eta_t)\delta_t + B\eta_t^2, \quad \text{if } t+1 \notin \mathcal{T}_T, \quad (26)$$

and

$$\delta_{t+1} \leq (1 - C\eta_t)\delta_{t+1-\tau} + B\eta_t^2, \quad \text{if } t+1 \in \mathcal{T}_T, \quad (27)$$

where

$$B := 6L\psi + \sum_{k \in \mathcal{K}} p_k^2 (N(\text{Var}(f) + M^2) + \theta_k^2) + 8(\tau-1)^2 \sum_{k \in \mathcal{K}} p_k \theta_k^2. \quad (28)$$

By setting the learning rate  $\eta_t$  and the FL system parameters, combined with some smoothness assumption of the local objective functions, we can prove (17).

Specifically, we set the diminishing learning rate  $\eta_t := \frac{\zeta}{t+\alpha}$  for some  $\zeta > 0$  and  $\alpha \geq \max\{4L\zeta, \tau\}$  such that  $\eta_t \leq \frac{1}{4L}$  and  $\eta_t \leq 2\eta_{t+\tau}$  to ensure the validity of Lemmas 12 and 13. Under this setup, we will demonstrate the existence of a finite  $\nu, \alpha > 0$  such that

$$\delta_t \leq \frac{\nu}{t+\alpha}$$

for all  $t \in \mathcal{T}$ . We prove it by induction. For  $t = 0$ , the above holds if  $\nu \geq \alpha\delta_0$ . Assuming that the above holds for some  $t = m\tau$ , where  $m \in \mathbb{Z}_+$ , it follows by inductively applying (26) that

$$\begin{aligned} \delta_{t+\tau} &\leq \delta_t \prod_{i=t}^{t+\tau-1} (1 - C\eta_i) + B \underbrace{\sum_{i=1}^{t+\tau-1} \eta_i^2}_{\leq 1} \cdot \prod_{i=t}^{t+\tau-1} (1 - C\eta_i) \\ &\leq \delta_t \prod_{i=t}^{t+\tau-1} (1 - C\eta_i) + B \sum_{i=1}^{t+\tau-1} \eta_i^2 \\ &\leq \delta_t \prod_{i=1}^{t+\tau-1} (1 - C\eta_i) + 4B\tau\eta_t^2 \\ &\leq \delta_t \exp\left(-C \sum_{i=1}^{t+\tau-1} \eta_i\right) + 4B\tau\eta_t^2 \\ &= \delta_t \exp\left(-C \sum_{i=1}^{t+\tau-1} \frac{\zeta}{i+\alpha}\right) + 4B\tau\eta_t^2 \\ &\leq \delta_t \exp\left(-\frac{C\zeta\tau}{t+\tau+\alpha}\right) + 4B\tau\eta_t^2 \\ &\leq \delta_t \left(1 - \frac{C\zeta\tau}{2(t+\tau+\alpha)}\right) + \frac{4B\tau\zeta^2}{(t+\alpha)^2}, \end{aligned}$$

where we used the inequality  $e^{-x} \leq 1 - x/2$  for  $x \in [0, 1]$ . Using our induction hypothesis that  $\delta_{m\tau} \leq \frac{\nu}{m\tau+\alpha}$ ,

$$\begin{aligned} \delta_{(m+1)\tau} &= \delta_{m\tau+\tau} \\ &\leq \frac{\nu}{m\tau+\alpha} \delta_t \left(1 - \frac{C\zeta\tau}{2(t+\tau+\alpha)}\right) + \frac{4B\tau\zeta^2}{(t+\alpha)^2} \\ &= \frac{\nu}{m\tau+\alpha} - \frac{\nu}{(m+1)\tau+\alpha} - \frac{\nu C\zeta\tau}{2(m\tau+\alpha)((m+1)\tau+\alpha)} + \frac{4B\tau\zeta^2}{(m\tau+\alpha)^2} + \frac{\nu}{(m+1)\tau+\alpha} \\ &= \frac{\nu\tau}{(m\tau+\alpha)((m+1)\tau+\alpha)} \left(1 - \frac{C\zeta}{2}\right) + \frac{4B\tau\zeta^2}{(m\tau+\alpha)^2} + \frac{\nu}{(m+1)\tau+\alpha}. \end{aligned}$$

Therefore,  $\delta_{(m+1)\tau} \leq \frac{\nu}{(m+1)\tau+\alpha}$  holds when

$$\begin{aligned} \frac{4B\tau\zeta^2}{(m\tau+\alpha)^2} &\leq \frac{\nu\tau}{(m\tau+\alpha)((m+1)\tau+\alpha)} \left(\frac{C\zeta}{2} - 1\right) \\ \frac{4B\zeta^2}{m\tau+\alpha} &\leq \frac{\nu}{(m+1)\tau+\alpha} \left(\frac{C\zeta}{2} - 1\right) \\ \nu &\geq \frac{(m+1)\tau+\alpha}{m\tau+\alpha} \cdot \frac{4B\zeta^2}{\frac{C\zeta}{2} - 1} \end{aligned} \quad (29)$$

Note that for  $m \geq 0$ ,

$$\frac{m\tau+\tau+\alpha}{m\tau+\alpha} \geq \frac{\tau+\alpha}{\alpha},$$

it suffices to choose  $\nu \geq \frac{4B\zeta^2(\tau+\alpha)}{\alpha(\frac{C\zeta}{2}-1)}$ .

Finally, by the smoothness property of the local objective functions

$$\mathbb{E}[F(\mathbf{W}_T)] - F(\mathbf{w}^*) \leq \frac{L}{2}\delta_T \leq \frac{L\nu}{2(T+\alpha)}, \quad (30)$$

provided that  $\nu \geq \max \left\{ \alpha\delta_0, \frac{4B\zeta^2(\tau+\alpha)}{\alpha(\frac{C\zeta}{2}-1)} \right\}$ ,  $\alpha \geq \{4L\zeta, \tau\}$  and  $\zeta > 0$ . In particular, setting  $\zeta = \frac{\tau}{C}$  implies that  $\alpha \geq \tau \max \left\{ \frac{4L}{C}, 1 \right\}$  and  $\nu \geq \max \left\{ \frac{4B\tau^2(\tau+\alpha)}{C^2\alpha(\tau-1)}, \alpha\|\mathbf{W}_0 - \mathbf{w}^*\| \right\}$ .

We emphasize that the above analysis is for synchronous communication rounds  $t \in \mathcal{T}$ . A decreasing upper bound for general  $t \geq 0$  can be obtained in a similar manner by utilizing standard gradient-descent analysis.

#### APPENDIX D PROOF OF LEMMA 12

Note that

$$\begin{aligned} \bar{\mathbf{J}}_t - \mathbf{J}_t &= \sum_{k \in \mathcal{K}} p_k \nabla F_k(\mathbf{W}_t^k) - p_k \left( \nabla F_k^{j_t^k}(\mathbf{W}_t^k) - \mathbf{E}_t^k \right) \\ &= \sum_{k \in \mathcal{K}} p_k (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)) + \sum_{k \in \mathcal{K}} p_k \mathbf{E}_t^k. \end{aligned}$$

Taking squared norm and expectation on both sides yield

$$\begin{aligned} \mathbb{E}[\|\bar{\mathbf{J}}_t - \mathbf{J}_t\|^2] &= \mathbb{E} \left[ \left\| \sum_{k \in \mathcal{K}} p_k (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)) + \sum_{k \in \mathcal{K}} p_k \mathbf{E}_t^k \right\|^2 \right] \\ &= \mathbb{E} \left[ \left\| \sum_{k \in \mathcal{K}} p_k (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)) \right\|^2 \right] + \mathbb{E} \left[ \left\| \sum_{k \in \mathcal{K}} p_k \mathbf{E}_t^k \right\|^2 \right] \\ &\quad + 2 \underbrace{\mathbb{E} \left\langle \sum_{k \in \mathcal{K}} p_k (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)), \sum_{k' \in \mathcal{K}} p_{k'} \mathbf{E}_t^{k'} \right\rangle}_{=0 \text{ as } \mathbb{E}(\mathbf{E}_t^k)=0} \\ &\leq \mathbb{E} \left[ \sum_{k \in \mathcal{K}} \left\| p_k (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)) \right\|^2 \right] + \mathbb{E} \left[ \sum_{k \in \mathcal{K}} \|p_k \mathbf{E}_t^k\|^2 \right] \\ &= \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \left\| (\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)) \right\|^2 \right] + \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \|\mathbf{E}_t^k\|^2 \right] \\ &\stackrel{(a)}{\leq} \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \left\| \nabla F_k^{j_t^k}(\mathbf{W}_t^k) \right\|^2 \right] + \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \|\mathbf{E}_t^k\|^2 \right] \\ &\stackrel{(b)}{=} \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \left\| \nabla F_k^{j_t^k}(\mathbf{W}_t^k) \right\|^2 \right] + \sum_{k \in \mathcal{K}} p_k^2 \mathbb{E} \left[ \|\mathbf{Z}_{t-1}^k + \mathbf{C}_{t-1}^k\|^2 \right] \\ &\stackrel{(c)}{\leq} \sum_{k \in \mathcal{K}} p_k^2 \theta_k^2 + \sum_{k \in \mathcal{K}} p_k^2 N(\text{Var}(f) + M^2) \\ &= \sum_{k \in \mathcal{K}} p_k^2 (\theta_k^2 + N(\text{Var}(f) + M^2)), \end{aligned}$$

where (a) holds since  $\mathbb{E}[\nabla F_k^{j_t^k}(\mathbf{W})] = \nabla F_k(\mathbf{W})$ , implying  $\mathbb{E}[\|\nabla F_k(\mathbf{W}_t^k) - \nabla F_k^{j_t^k}(\mathbf{W}_t^k)\|^2] \leq \mathbb{E}[\|\nabla F_k^{j_t^k}(\mathbf{W}_t^k)\|^2]$ , (b) is due to norm clipping, (c) holds by **AS2** and  $\mathbb{E}[\|\mathbf{Z}_t^k\|^2] \in \{0, N\text{Var}(f)\}$ , hence  $\mathbb{E}[\|\mathbf{E}_t^k\|^2] \leq N(\text{Var}(\hat{f}) + M^2)$  for all  $t$ .

APPENDIX E  
PROOF OF LEMMA 13

Since FedAvg contains  $\tau$  steps in one communication round, it follows that for  $t \geq 0$ , there exists a  $t_0 \leq t$  such that  $t_0 \in \mathcal{T}_T$ ,  $t - t_0 \leq \tau - 1$  and  $\mathbf{W}_{t_0}^k = \mathbf{W}'_{t_0}$  for all  $k \in \mathcal{K}$ . Note that (25) holds trivially for  $t = t_0$ . For  $t > t_0$ , we have

$$\begin{aligned} \mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|\mathbf{W}'_t - \mathbf{W}_t^k\|^2 \right] &= \mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|(\mathbf{W}_t^k - \mathbf{W}'_{t_0}) - (\mathbf{W}'_t - \mathbf{W}'_{t_0})\|^2 \right] \\ &\stackrel{(a)}{\leq} \mathbb{E} \left[ \sum_{k \in \mathcal{K}} p_k \|\mathbf{W}_t^k - \mathbf{W}'_{t_0}\|^2 \right] \\ &= \sum_{k \in \mathcal{K}} p_k \mathbb{E} [\|\mathbf{W}_t^k - \mathbf{W}'_{t_0}\|^2] \\ &\stackrel{(b)}{=} \sum_{k \in \mathcal{K}} p_k \mathbb{E} \left[ \left\| \sum_{t'=t_0}^{t-1} \eta_{t'} \nabla F_k^{j_{t'}}(\mathbf{W}_{t'}^k) \right\|^2 \right] \\ &\stackrel{(c)}{\leq} \sum_{k \in \mathcal{K}} p_k (\tau - 1) \sum_{t'=t_0}^{t-1} \eta_{t'}^2 \mathbb{E} [\|\nabla F_k^{j_{t'}}(\mathbf{W}_{t'}^k)\|^2] \\ &\stackrel{(d)}{\leq} \sum_{k \in \mathcal{K}} p_k (\tau - 1)^2 \eta_{t_0}^2 \theta_k^2 \\ &\stackrel{(e)}{\leq} 4(\tau - 1)^2 \eta_t^2 \sum_{k \in \mathcal{K}} p_k \theta_k^2 \end{aligned}$$

where (a) holds since  $\mathbb{E}[\|X - \mathbb{E}[X]\|^2] \leq \mathbb{E}[\|X\|^2]$  where  $X := \mathbf{W}_t^k - \mathbf{W}'_{t_0}$ , (b) holds since  $\mathbf{E}_{t'}^k = \mathbf{0}$  for  $t' = t_0, \dots, t-1$  and iterates recursively over the first case of (18), (c) holds since  $\|\sum_{t'=t_0}^{t-1} \mathbf{r}_{t'}\|^2 \leq (\tau - 1 - t_0) \sum_{t'=t_0}^{t-1} \|\mathbf{r}_{t'}\|^2 \leq (\tau - 1) \sum_{t'=t_0}^{t-1} \|\mathbf{r}_{t'}\|^2$ , (d) holds by Assumption 2, and (e) holds due to  $\eta_{t_0} \leq \eta_{t-\tau} \leq 2\eta_t$ .

## REFERENCES

- [1] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, “Communication-efficient learning of deep networks from decentralized data,” in *AISTATS*, 2016.
- [2] J. Dean, G. Corrado, R. Monga, K. Chen, M. Devin, M. Mao, M. a. Ranzato, A. Senior, P. Tucker, K. Yang, Q. Le, and A. Ng, “Large scale distributed deep networks,” in *NeurIPS*, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, Eds., vol. 25, 2012.
- [3] T. Li, A. Sahu, A. Talwalkar, and V. Smith, “Federated learning: Challenges, methods, and future directions,” *arXiv preprint arXiv:1908.07873*, 08 2019.
- [4] J. Konečný, H. B. McMahan, F. X. Yu, P. Richtarik, A. T. Suresh, and D. Bacon, “Federated learning: Strategies for improving communication efficiency,” in *NIPS Workshop on Private Multi-Party Machine Learning*, 2016.
- [5] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, “Deep gradient compression: Reducing the communication bandwidth for distributed training,” *arXiv preprint arXiv:1712.01887*, 2020.
- [6] C. Hardy, E. Le Merrer, and B. Sericola, “Distributed deep learning on edge-devices: Feasibility via adaptive compression,” in *IEEE NCA*, 2017, pp. 1–8.
- [7] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, “Qsgd: Communication-efficient sgd via gradient quantization and encoding,” in *NeurIPS*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30, 2017.
- [8] N. Shlezinger, M. Chen, Y. C. Eldar, H. V. Poor, and S. Cui, “UVeQFed: Universal vector quantization for federated learning,” *IEEE Trans. Signal Process*, vol. 69, pp. 500–514, 2020.
- [9] C. Dwork, F. McSherry, K. Nissim, and A. Smith, “Calibrating noise to sensitivity in private data analysis,” in *Theory of Cryptography*. Springer Berlin Heidelberg, 2006, pp. 265–284.
- [10] C. Dwork and A. Roth, “The algorithmic foundations of differential privacy,” *Found. Trends Theor. Comput. Sci.*, vol. 9, no. 3-4, pp. 211–407, 2014.
- [11] A. Girgis, D. Data, S. Diggavi, P. Kairouz, and A. T. Suresh, “Shuffled model of differential privacy in federated learning,” in *AISTATS*. PMLR, 2021, pp. 2521–2529.
- [12] N. Agarwal, P. Kairouz, and Z. Liu, “The skellam mechanism for differentially private federated learning,” in *NeurIPS*, ser. NIPS ’21, Red Hook, NY, USA, 2024.
- [13] L. Roberts, “Picture coding using pseudo-random noise,” *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 145–154, 1962.
- [14] L. Schuchman, “Dither signals and their effect on quantization noise,” *IEEE Transactions on Communication Technology*, vol. 12, no. 4, pp. 162–165, 1964.
- [15] J. O. Limb, “Design of dither waveforms for quantized visual signals,” *The Bell System Technical Journal*, vol. 48, no. 7, pp. 2555–2582, 1969.
- [16] N. S. Jayant and L. R. Rabiner, “The application of dither to the quantization of speech signals,” *The Bell System Technical Journal*, vol. 51, no. 6, pp. 1293–1304, 1972.
- [17] A. Sripad and D. Snyder, “A necessary and sufficient condition for quantization errors to be uniform and white,” *IEEE Trans. Acoust., Speech, Signal Process*, vol. 25, no. 5, pp. 442–448, 1977.
- [18] M. Kim, O. Günlü, and R. F. Schaefer, “Federated learning with local differential privacy: Trade-offs between privacy, utility, and communication,” *ICASSP*, pp. 2650–2654, 2021.
- [19] A. M. Shahmiri, C. W. Ling, and C. T. Li, “Communication-efficient Laplace mechanism for differential privacy via random quantization,” in *ICASSP*. IEEE, 2024, pp. 4550–4554.
- [20] B. Hasircioğlu and D. Gündüz, “Communication efficient private federated learning using dithering,” in *ICASSP*, 2024, pp. 7575–7579.
- [21] G. Yan, T. Li, T. Lan, K. Wu, and L. Song, “Layered randomized quantization for communication-efficient and privacy-preserving distributed learning,” *arXiv preprint arXiv:2312.07060*, 2023.

- [22] M. Hegazy, R. Leluc, C. T. Li, and A. Dieuleveut, "Compression with exact error distribution for federated learning," in *AISTATS*, vol. 238, 2024, pp. 613–621.
- [23] N. Lang, E. Sofer, T. Shaked, and N. Shlezinger, "Joint privacy enhancement and quantization in federated learning," *IEEE Trans. Signal Process*, vol. 71, pp. 295–310, 2023.
- [24] C. H. Bennett, P. W. Shor, J. Smolin, and A. V. Thapliyal, "Entanglement-assisted capacity of a quantum channel and the reverse Shannon theorem," *IEEE Trans. Inf. Theory*, vol. 48, no. 10, pp. 2637–2655, 2002.
- [25] P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan, "The communication complexity of correlation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 438–449, Jan 2010.
- [26] C. T. Li and A. El Gamal, "Strong functional representation lemma and applications to coding theorems," *IEEE Trans. Inf. Theory*, vol. 64, no. 11, pp. 6967–6978, Nov 2018.
- [27] M. Havasi, R. Pecharz, and J. M. Hernández-Lobato, "Minimal random code learning: Getting bits back from compressed model parameters," in *7th Int. Conf. Learn. Represent.*, 2019.
- [28] A. Shah, W.-N. Chen, J. Balle, P. Kairouz, and L. Theis, "Optimal compression of locally differentially private mechanisms," in *AISTATS*. PMLR, 2022, pp. 7680–7723.
- [29] G. Flamich, "Greedy Poisson rejection sampling," in *NeurIPS*, 2023.
- [30] Y. Liu, W.-N. Chen, A. Özgür, and C. T. Li, "Universal exact compression of differentially private mechanisms," in *NeurIPS 2024*, 2024.
- [31] C. T. Li, *Channel Simulation: Theory and Applications to Lossy Compression and Differential Privacy*. Hanover, MA, USA: Now Publishers Inc., Dec. 2024, vol. 21, no. 6.
- [32] S. Walker, "The uniform power distribution," *Journal of Applied Statistics*, vol. 26, no. 4, pp. 509–517, 1999.
- [33] S. Choy and S. G. Walker, "The extended exponential power distribution and bayesian robustness," *Statistics & Probability Letters*, vol. 65, no. 3, pp. 227–232, 2003.
- [34] E. Agustsson and L. Theis, "Universally quantized neural compression," *NeurIPS*, vol. 33, pp. 12 367–12 376, 2020.
- [35] M. Hegazy and C. T. Li, "Randomized quantization with exact error distribution," in *IEEE ITW*. IEEE, 2022, pp. 350–355.
- [36] C. W. Ling and C. T. Li, "Rejection-sampled universal quantization for smaller quantization errors," *IEEE Trans. Inf. Theory*, vol. 71, no. 12, pp. 9784–9803, 2025.
- [37] C. W. Ling, C. H. M. Shiu, Y. Wu, J. Sun, C. T. Li, L. Song, and W. Xu, "Communication-efficient and privacy-adaptable mechanism for federated learning," *arXiv preprint arXiv:2501.12046*, 2025.
- [38] S. U. Stich, "Local SGD converges fast and communicates little," in *International Conference on Learning Representations*, 2019.
- [39] J. Ziv, "On universal quantization," *IEEE Trans. Inf. Theory*, vol. 31, no. 3, pp. 344–347, 1985.
- [40] R. Zamir and M. Feder, "On universal quantization by randomized uniform/lattice quantizers," *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 428–436, 1992.
- [41] J. H. Conway and N. J. A. Sloane, *Sphere packings, lattices and groups*. Springer Science & Business Media, 2013, vol. 290.
- [42] R. Zamir, *Lattice Coding for Signals and Networks*. Cambridge University Press, 2014.
- [43] R. M. Gray and T. G. Stockham, "Dithered quantizers," *IEEE Trans. Inf. Theory*, vol. 39, no. 3, pp. 805–812, 1993.
- [44] A. Kirac and P. Vaidyanathan, "Results on lattice vector quantization with dithering," *IEEE Transactions On Circuits and Systems II: Analog and Digital Signal Processing*, vol. 43, no. 12, pp. 811–826, 1996.
- [45] R. Zamir and M. Feder, "On lattice quantization noise," *IEEE Trans. Inf. Theory*, vol. 42, no. 4, pp. 1152–1159, 1996.
- [46] D. B. Wilson, "Layered multishift coupling for use in perfect sampling algorithms (with a primer on CFTP)," *Monte Carlo Methods*, vol. 26, pp. 141–176, 2000.
- [47] C. P. Robert and G. Casella, "Monte Carlo statistical methods," *Springer Texts in Statistics*, p. 274, 2004.
- [48] L. Zhu, Z. Liu, and S. Han, *Deep leakage from gradients*. Red Hook, NY, USA: Curran Associates Inc., 2019.
- [49] B. Zhao, K. R. Mopuri, and H. Bilen, "idlq: Improved deep leakage from gradients," *arXiv preprint arXiv:2001.02610*, 2020.
- [50] Y. Huang, S. Gupta, Z. Song, K. Li, and S. Arora, "Evaluating gradient inversion attacks and defenses in federated learning," in *NeurIPS*, ser. NIPS '21. Red Hook, NY, USA: Curran Associates Inc., 2021.
- [51] B. Balle, G. Barthe, and M. Gaboardi, "Privacy amplification by subsampling: Tight analyses via couplings and divergences," in *NeurIPS*, S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, Eds., vol. 31, 2018.
- [52] J. Dong, A. Roth, and W. J. Su, "Gaussian differential privacy," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, vol. 84, no. 1, pp. 3–37, 02 2022.
- [53] I. Mironov, "Rényi differential privacy," in *2017 IEEE 30th Computer Security Foundations Symposium (CSF)*. Los Alamitos, CA, USA: IEEE Computer Society, Aug. 2017, pp. 263–275.
- [54] S. Golomb, "Run-length encodings (corresp.)," *IEEE Trans. Inf. Theory*, vol. 12, no. 3, pp. 399–401, 1966.
- [55] R. Gallager and D. van Voorhis, "Optimal source codes for geometrically distributed integer alphabets (corresp.)," *IEEE Trans. Inf. Theory*, vol. 21, no. 2, pp. 228–230, 1975.
- [56] X. Li, K. Huang, W. Yang, S. Wang, and Z. Zhang, "On the convergence of fedavg on non-iid data," in *International Conference on Learning Representations*, 2020.
- [57] G. An, "The effects of adding noise during backpropagation training on a generalization performance," *Neural Computation*, vol. 8, no. 3, pp. 643–674, 1996.