On Differential Privacy for Wireless Federated Learning with Non-coherent Aggregation

Mohamed Seif* Alphan Şahin† H. Vincent Poor* Andrea J. Goldsmith*
*Princeton University, Princeton, NJ, USA and †University of South Carolina, Columbia, SC, USA
E-mails: {*mseif, poor, goldsmith*}@princeton.edu, *asahin*@mailbox.sc.edu

***Abstract*—In this paper, we study distributed training by majority vote with the sign stochastic gradient descent (signSGD) along with over-the-air computation (OAC) under local differential privacy constraints. In our approach, the users first clip the local stochastic gradients and inject a certain amount of noise as a privacy enhancement strategy. Subsequently, they activate the indices of OFDM subcarriers based on the signs of the perturbed local stochastic gradients to realize a frequency-shift-keying-based majority vote computation at the parameter server. We evaluate the privacy benefits of the proposed approach and characterize the per-user privacy leakage theoretically. Our results show that the proposed technique improves the privacy guarantees and limits the leakage to a scaling factor of $\mathcal{O}(1/\sqrt{K})$, where K is the number of users, thanks to the superposition property of the wireless channel. With numerical experiments, we show that the proposed non-coherent aggregation is superior to quadrature-phase-shift-keying-based coherent aggregation, namely, one-bit digital aggregation (OBDA), in learning accuracy under time synchronization errors when the same privacy enhancement strategy is introduced to both methods.**

***Index Terms*—Federated learning over wireless networks, majority vote, differential privacy, noise injection, over-the-air computation.**

I. Introduction

Wireless federated learning (FL) is a distributed learning paradigm where many users communicate with a parameter server (PS) to train a neural network over a wireless network. To address the spectrum congestion caused by a large number of users participating in learning, recent studies, e.g., [1]–[4], have investigated several over-the-air computation (OAC) schemes for aggregating local information in the communication channel. These schemes often require the users' signals to be received with similar amplitudes for coherent superposition. Hence, they rely on precise time synchronization among users and the availability of accurate channel state information (CSI) at the users for pre-equalization. Practical time synchronization mechanisms, such as timing advance in 5G NR [5], can achieve synchronization within a duration, e.g., a cyclic prefix (CP) range of an orthogonal frequency division multiplexing (OFDM) symbol. However, precise sample-level time synchronization is not trivial to maintain due to the synchronization impairments at the users and PS. Furthermore, the phase rotation due to time synchronization errors alter CSI and distorts the coherent superposition [6].

Most of the OAC schemes for wireless FL in the literature adopt pre-equalization techniques to address the impact of multipath distortion on the transmitted symbols. For instance, Zhu et al. [1] investigate analog modulation over OFDM for broadband analog aggregation (BAA) in which OFDM subcarriers are modulated with model parameters at the users. The symbols on the OFDM subcarriers are multiplied by the inverse of the channel coefficients and faded subcarriers are excluded from transmission, i.e., truncated-channel inversion (TCI). Similarly, [7] proposes one-bit broadband digital aggregation (OBDA) with TCI. In this method, the users utilize quadrature phase-shift keying (QPSK) symbols over OFDM subcarriers, where the real and imaginary parts of QPSK symbols are formed by using the signs of the stochastic gradients. This transmission scheme can be viewed as QPSK along with majority voting (namely, QPSK-MV) detected at the PS. To eliminate the need for CSI at the users, in [2] and [8], blind users are considered, and an aggregated CSI is used at the PS along with multiple antennas for computation. Although this approach addresses the computation problem under fading channels without using CSI at the users, it requires a large number of antennas to achieve channel hardening. To eliminate the need for CSI at the users and PS and phase synchronization, another approach is to use non-coherent methods. In [9], Goldenbaum and Stanczak introduce an approach that relies on modulating a sequence with the continuous-valued parameters and calculating the energy of the superposed signal at the PS. In [10] and [11], the authors realize distributed training by the majority vote (MV) with sign stochastic gradient descent (signSGD) [12] over a wireless network by calculating the MV based on orthogonal signaling at the users and a non-coherent detector at the PS, where the modulations used in these papers are frequency-shift keying (FSK), pulse-position modulation (PPM), and chirp-shift keying (CSK). Among these schemes, FSK is shown to be most spectrally-efficient as it does not require a guard time to accommodate the time-synchronization errors. It is also demonstrated that it can work in practice [13]. Due to its robustness against imperfections, in this work, we adopt FSK-based MV (FSK-MV) for wireless FL and expand it with a new privacy amplification method.

Recently, much attention has been given to developing differentially private FL methods, and differential privacy (DP), as introduced by Dwork et al. [14], has emerged as the standard approach to private data analysis and aggregation. In the context of FL, local differential privacy (LDP) is more

This work has been supported in part by U.S National Science Foundation under Grants CCF-1908308, CNS-2128448, CNS-2147631, and the AFOSR Grant Award no. 002484665.

suitable since it enables users to locally perturb and disclose data to an untrusted data curator or aggregator [15]. Several studies design FL algorithms that satisfy LDP [16], [17], which typically requires considerable perturbation noise to ensure privacy guarantees. Differential privacy in wireless FL has been studied in several works [4], [18]–[20]. In particular, Seif et al. [18] showed that the superposition nature of the wireless channel provides boosts levels of LDP guarantees due to noise amplification that the users add. It is worth highlighting that CSI is assumed to be perfect in that work, which is crucial to align the users' gradients at the PS and amplify the net received noise seen at the PS. In this paper, we are interested in answering the following two fundamental questions: (1) *Can we achieve provable convergence and local differential privacy guarantees for wireless federated learning by just sending the signs of the local gradients without relying on the CSI of the users? If yes,* (2) *can we still achieve the same privacy guarantees compared to the case with perfect CSI scenario?*

Main contributions: We answer the above two questions in the affirmative by presenting a novel differentially private non-coherent transmission scheme for distributed signSGD that utilizes OFDM subcarriers with FSK modulation. Furthermore, we derive the convergence rate to a stationary point for general non-convex loss functions. In addition, we formally characterize the privacy guarantees as a function of the wireless channel parameters. Interestingly, we show that the privacy leakage still *scales* as $\mathcal{O}(1/\sqrt{K})$, i.e., the same privacy guarantees in the perfect CSI case as shown in [18]. To the best of our knowledge, this is the first result on wireless FL under non-coherent transmission and LDP constraints.

II. Problem Statement & System Model

A. Wireless channel model

We consider a single-antenna wireless FL system with K users and a central PS, as shown in Fig. 1. The users are connected to the PS through a wireless fading multiple-access channel (MAC). We assume that the users access the channel with OFDM symbols and the averaged received signal powers of the users are identical at the PS's location. Assuming that the CP duration is larger than the sum of the maximum-excess delay of the multi-path channel and the maximum time synchronization error, the input-output relationship at the tth communication round can be expressed as

$$y_s^{(t)} = \sum_{k=1}^{K} h_{k,s}^{(t)} x_{k,s}^{(t)} + m_s^{(t)}, \forall s \in \{1, 2, \cdots, S\}, \qquad (1)$$

where $x_{k,s}^{(t)}$ is the sth transmitted symbol of user k at the tth communication round, $y_s^{(t)}$ is the received symbol at the PS, $h_{k,s}^{(t)} \in \mathbb{C}$ is the normalized Rayleigh fading channel coefficient between the user k and the PS, $m_s^{(t)} \in \mathbb{C}$ is the receiver noise with a zero-mean symmetric complex Gaussian distribution with variance σ_m^2, and S is the total number of complex dimensions used for the transmissions.

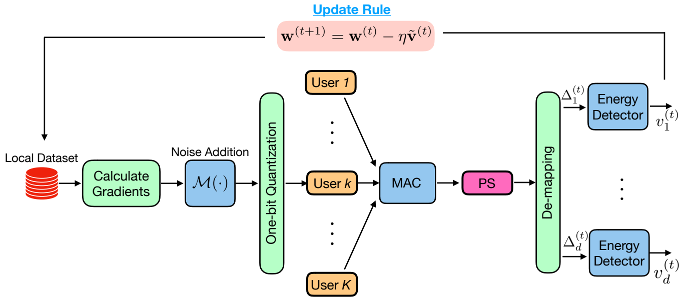

Fig. 1. Illustration of the distributed training based on signSGD framework with noise injection. The users collaborate with the PS to jointly train a machine learning model over a fading MAC. The interaction between users and the PS must satisfy local differential privacy.

B. Federated learning model

Each user k has a private local dataset $\mathcal{D}_k$ with D_k data points, denoted as $\mathcal{D}_k = \{(\mathbf{u}_i^{(k)}, v_i^{(k)})\}_{i=1}^{D_k}$, where $\mathbf{u}_i^{(k)}$ is the ith data point and $v_i^{(k)}$ is the corresponding label at the kth user. The local loss function at the kth user is given by

$$f_k(\mathbf{w}) = \frac{1}{D_k} \sum_{i=1}^{D_k} f(\mathbf{w}; (\mathbf{u}_i^{(k)}, v_i^{(k)})),$$

where $\mathbf{w} \in \mathbb{R}^d$ is the parameter vector to be optimized. Users communicate with the PS through the fading MAC as described above to train a model by minimizing the loss function $F(\mathbf{w})$, i.e.,

$$\mathbf{w}^* = \arg\min_{\mathbf{w}} F(\mathbf{w}) \triangleq \frac{1}{\sum_{k=1}^{K} D_k} \sum_{k=1}^{K} D_k f_k(\mathbf{w}).$$

The minimization of $F(\mathbf{w})$ is carried out iteratively through distributed training by MV with the signSGD algorithm. In particular, for the tth training iteration, the PS broadcasts the global parameter vector $\mathbf{w}^{(t)}$ to all users. The user k computes its local gradient using stochastic mini batch $\mathcal{B}_k \subseteq \mathcal{D}_k$, with size n_b (i.e., $|\mathcal{B}_k| = n_b$) as

$$\mathbf{g}_k^{(t)} = \frac{1}{n_b} \sum_{i \in \mathcal{B}_k} \nabla f_k(\mathbf{w}^{(t)}; (\mathbf{u}_i^{(k)}, v_i^{(k)})) \ ,$$

where $\mathbf{g}_k^{(t)}$ is the stochastic gradient vector of the user k. Instead of sending the actual values of the local gradients, the users then send the signs of their stochastic gradients to the PS. To this end, user k extracts the sign of each element of the computed stochastic gradient $\mathbf{g}_k^{(t)}$, i.e., $\tilde{g}_{k,i}^{(t)} \triangleq \text{sign}(g_{k,i}^{(t)})$. The PS obtains the MV for the ith gradient as follows $\tilde{v}_i^{(t)} \triangleq \text{sign}\left(\sum_{k=1}^{K} \tilde{g}_{k,i}^{(t)}\right)$. In this study, we define $\text{sign}(\cdot)$ as an operator that results in 1, -1, or ± 1 at random for a positive, a negative, or a zero-valued argument, respectively.

Subsequently, the global parameter $\mathbf{w}^{(t)}$ is updated using the sign vector $\tilde{\mathbf{v}}^{(t)} = (\tilde{v}_1^{(t)}, \tilde{v}_2^{(t)}, \cdots, \tilde{v}_d^{(t)})^T$ according to $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\tilde{\mathbf{v}}^{(t)}$, where η is the learning rate of the distributed training. The iteration process continues until a specified number of training iterations/communications rounds.

In addition, the signSGD algorithm must satisfy LDP constraints for each user, as defined next.

Definition 1. *((ϵ,δ)-LDP [14]) For a user k, a randomized mechanism $\mathcal{M}_k : \mathcal{D}_k \to \mathbb{R}^d$ is (ϵ,δ) LDP if for any $x, x' \in \mathcal{D}_k$, and any measurable subset $\mathcal{S} \subseteq Range(\mathcal{M}_k)$, we have*

$$\Pr(\mathcal{M}_k(x) \in \mathcal{S}) \leq e^{\epsilon} \Pr(\mathcal{M}_k(x') \in \mathcal{S}) + \delta. \quad (2)$$

The setting when $\delta = 0$ is referred as pure ϵ-LDP.

III. Main Results & Discussions

In this section, we first present the proposed scheme. We then derive the local privacy level where each user perturbs its local gradient vector via Gaussian artificial noise and then extracts the signs of the perturbed gradient for transmission. Finally, we present the convergence rate of the private wireless federated signSGD algorithm. Due to space limitations, we omit the proofs in this paper.

A. Proposed scheme: FSK-based MV with perturbation

We consider OFDM-based OAC discussed as follows:
(1) **Local perturbation noise for privacy.** At the tth training iteration, each user k computes a noisy version of its local gradient update as

$$\tilde{\mathbf{g}}_k^{(t)} = \mathbf{g}_k^{(t)} + \mathbf{n}_k^{(t)},$$

where $\mathbf{n}_k^{(t)} \sim \mathcal{N}(0, \sigma_k^{(t)} \mathbf{I}_d)$ is the artificial noise for privacy. We further assume that the norms of gradient vectors are bounded by some constant $C \geq 0$, and normalize the gradient vector to C, i.e., $\mathbf{g}_k^{(t)} := \min\left(1, C/\|\mathbf{g}_k^{(t)}\|_2\right) \cdot \mathbf{g}_k^{(t)}$.
(2) **One-bit quantization and signal modulation.** Consequently, each user performs one-bit quantization by computing the sign of each element of the local stochastic gradient $\tilde{\mathbf{g}}_k^{(t)}$. Further, we allocate two *orthogonal resources* based on the sign of the gradient. Specifically, the sign of each element $\tilde{g}_{k,i}$ is modulated with FSK as [10]

$$x_{k,2i-1}^{(t)} = \begin{cases} \sqrt{E_{\mathrm{s}}}, & \mathrm{sign}(\tilde{g}_{k,i}) = 1 \\ 0, & \text{otherwise} \end{cases},$$

and

$$x_{k,2i}^{(t)} = \begin{cases} \sqrt{E_{\mathrm{s}}}, & \mathrm{sign}(\tilde{g}_{k,i}) = -1 \\ 0, & \text{otherwise} \end{cases},$$

where $x_{k,2i-1}^{(t)}$ and $x_{k,2i}^{(t)}$ denote the symbols modulating two adjacent OFDM subcarriers at the user k for the ith gradient at the tth communication round and $E_s = 2$ is an energy normalization factor for FSK. Hence, with the proposed scheme, the number of utilized wireless resources S is equal to $2d$.
(3) **Energy detection at the PS.** The received signal at the PS at the two adjacent sub-carriers of the OFDM symbol for the ith coordinate can be written as:

$$y_{2i-1}^{(t)} = \sum_{\substack{\forall k: \\ \mathrm{sign}(\tilde{g}_{k,i})=1}} h_{k,2i-1}^{(t)} x_{k,2i-1}^{(t)} + m_{2i-1}^{(t)},$$

$$y_{2i}^{(t)} = \sum_{\substack{\forall k: \\ \mathrm{sign}(\tilde{g}_{k,i})=-1}} h_{k,2i}^{(t)} x_{k,2i}^{(t)} + m_{2i}^{(t)}.$$

The PS subsequently calculates the operation given by

$$\Delta_i^{(t)} = |y_{2i-1}^{(t)}|^2 - |y_{2i}^{(t)}|^2, \ \forall i \in \{1, 2, \cdots, d\},$$

followed by $\mathrm{sign}(\cdot)$ operation for each coordinate i, i.e., $\tilde{v}_i^{(t)} = \mathrm{sign}(\Delta_i^{(t)})$, to obtain the MVs. The PS then updates the global model $\mathbf{w}^{(t)}$ according to $\mathbf{w}^{(t+1)} = \mathbf{w}^{(t)} - \eta\tilde{\mathbf{v}}^{(t)}$.

B. Local differential privacy analysis

We analyze the privacy level achieved by our proposed scheme that adds artificial noise perturbations to privatize its local data. We focus on analyzing the privacy leakage under an additive noise mechanism that is drawn from a Gaussian distribution. This well-known perturbation technique is called the Gaussian mechanism, and it provides rigorous privacy guarantees, defined as follows:

Definition 2. *(Gaussian mechanism [14]) Suppose a node wants to release a function $f(X)$ of an input X subject to (ϵ,δ)-LDP. The Gaussian release mechanism is defined as*

$$\mathcal{M}(X) \triangleq f(X) + \mathcal{N}(0, \sigma^2 \mathbf{I}_d).$$

If the sensitivity of the function is bounded by Δ_f, i.e., $\|f(x) - f(x')\|_2 \leq \Delta_f$, $\forall x, x'$, then for any $\delta \in (0,1]$, the Gaussian mechanism satisfies (ϵ,δ)-LDP, where

$$\epsilon = \frac{\Delta_f}{\sigma}\sqrt{2\log\frac{1.25}{\delta}}. \quad (3)$$

Privacy Model: We assume the PS is honest but curious. It is honest in the sense that it follows the procedure accordingly, but it might learn sensitive information about users' data. Therefore, the proposed wireless FL algorithm should satisfy LDP constraints for each user. Even though the PS follows the non-coherent communication scheme, we consider the worst-case privacy model where the PS can have *perfect* global CSI and reconstruct the full received signal at each communication round. Our privacy model is considered as *robust* to any side information that an adversary (i.e., the PS) can have. The privacy guarantee of the proposed algorithm is presented in the following theorem.

Theorem 1. *(The worst-case privacy guarantee) For a user k, the proposed transmission scheme achieves (ϵ_k, δ)-LDP per iteration, where*

$$\epsilon_k = \frac{2\gamma_k |h_{k,\max}| \sqrt{E_s} C \times \sqrt{2\log 1.25/\delta}}{\sqrt{E_s\left(\sum_{j=1}^{K} |h_{j,\min}|^2 (\gamma_j^2 \sigma_j^2 + \sigma_d^2)\right) + \sigma_m^2}}, \quad (4)$$

where $|h_{k,\max}| \triangleq \max_{i\in[2d]} |h_{k,i}|$ and $|h_{k,\min}|$ is defined similarly, $\gamma_k = \sqrt{2/\pi\sigma_k^2}$ is the Bussgang's coefficient for one-bit quantization, and σ_d^2 is the variance of distortion noise due to quantization.

Remark 1. *From the above result, we can observe the synergistic benefits of wireless aggregation for amplifying the privacy levels for each user. Specifically, the privacy leakage per user, ϵ_k behaves as $\mathcal{O}(1/\sqrt{K})$. Besides the wireless channel noise, we also notice that the privacy guarantee can* *be further improved when considering the one-bit quantization distortion. Harnessing and modeling the intrinsic randomness (i.e., the distribution of distortion) of quantization for providing rigorous privacy levels is of independent interest and left as a future work.*

Proof Sketch: The key challenge in the privacy analysis is that the sign operation performed at each user before wireless transmission of the gradients is *non-linear*. Therefore, it is not straightforward to (i) observe the impact of the wireless aggregation, and (ii) directly apply the existing results of the Gaussian mechanism [14]. To analyze the privacy levels that our proposed scheme achieves, we first need to approximate the non-linear sign operation as a linear relation using Bussgang's decomposition method [21] for one-bit quantization. Note that the quantization operation is performed on a Gaussian vector with mean $\mathbf{g}_k^{(t)}$ and covariance $\sigma^2\mathbf{I}$. Following standard analysis for the Gaussian mechanism, we can show that each coordinate of the estimated gradient at the PS is perturbed with different amount of noise σ_i^2 due to the impact of the *frequency-selective* fading channels. We then account for the worst case privacy per coordinate which corresponds to the minimum amount of perturbation noise. For more technical details about the proof, we refer the readers to [14].

In order to further show synergistic benefits of the wireless aggregation, we upper bound the achievable ϵ_k from Theorem 1 as follows.

Corollary 1. *(Privacy scaling) For a user k, the proposed transmission scheme achieves (ϵ_k, δ)-LDP per iteration, where*

$$\epsilon_k \leq \frac{|h_{k,\max}|}{\sqrt{\sum_{j=1}^{K}|h_{j,\min}|^2}} \cdot \frac{2C}{\sigma_k}\sqrt{2\log\frac{1.25}{\delta}}. \tag{5}$$

Interestingly, we can observe that the privacy parameter ϵ_k behaves as $\mathcal{O}(1/\sqrt{K})$ (see Fig. 2), which scales similarly as the perfect CSI scenario in [18]. We next analyze the convergence rate of our proposed wireless Federated signSGD for general non-convex loss function under the following assumptions as in [10], [12].

C. Federated learning convergence

Assumption 1. *(Smoothness) Let $\mathbf{g}(\mathbf{w})$ denote the true gradient of the global loss function $F(\mathbf{w})$ evaluated at $\mathbf{w}$. Then for any $\mathbf{w}$ and $\mathbf{w}'$, the loss function $F(\mathbf{w})$ is $\mathbf{L}$-smooth if*

$$|F(\mathbf{w}) - F(\mathbf{w}') - \mathbf{g}(\mathbf{w})^T(\mathbf{w}-\mathbf{w}')| \leq \frac{1}{2}\sum_{i=1}^{d} L_i(\mathbf{w}_i' - \mathbf{w}_i),$$

holds for a non-negative constant vector $\mathbf{L} = (L_1, L_2, \cdots, L_d)^T$.

Assumption 2. *(Bounded variance) The local stochastic gradient estimate of user k has a coordinate bounded variance, i.e., $\mathbb{E}\left[(g_{k,i}^{(t)} - g_i^{(t)})^2\right] \leq \sigma_i^2/n_b$, that holds for a non-negative constant vector $\sigma = (\sigma_1, \sigma_2, \cdots, \sigma_d)^T$.*

Assumption 3. *(Unimodal symmetry) Given any model $\mathbf{w}^{(t)}$, each coordinate of the stochastic gradient estimate $\tilde{\mathbf{g}}_{k,i}, \forall k, i$ has a unimodal distribution that is also symmetric around its mean.*

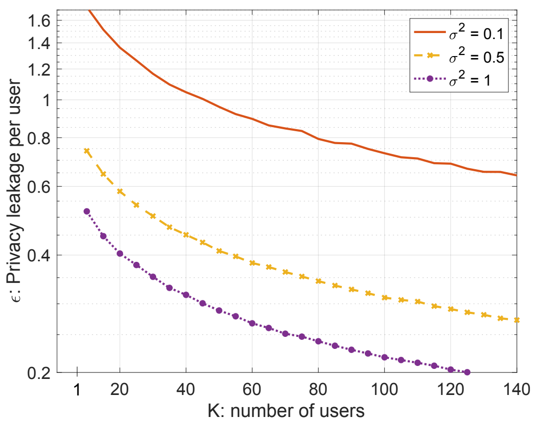

Fig. 2. Privacy Scaling: Comparison for local privacy leakage per user presented in eqn. (5) as a function of K for different values of local perturbation noises σ's, where $E_s = 2$, $C = 0.1$ and $\delta = 10^{-5}$.

We are now ready to present our convergence result as a function of the wireless channel, local perturbation noises and transmit powers in the following theorem.

Theorem 2. *(Utility guarantee) Suppose the global loss function $F(\mathbf{w})$ satisfies the above assumptions. Then, for some constant $c > 0$, a number of iterations T, a batch size $n_b = T/c$, and a learning rate $\eta = 1/\sqrt{\|\mathbf{L}\|_1 n_b}$, the convergence rate of the private wireless FL algorithm is*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\mathbf{g}^{(t)}\|_1\right] \leq \sqrt{\frac{\|\mathbf{L}\|_1}{Tc}}\left(1+\frac{2}{K\cdot\mathrm{SNR}}\right)(R+c/2) + \frac{\sqrt{8}c}{3\sqrt{T}}\|\sigma\|_1, \tag{6}$$

where $R \triangleq F(\mathbf{w}^{(0)}) - F(\mathbf{w}^*)$, $\mathrm{SNR} \triangleq E_s/\sigma_m^2$, and σ is the vector containing the variances of the effective perturbation per coordinate due to data subsampling and local perturbation for privacy.

Remark 2. *It is worth noting that the local perturbation noise via the Gaussian mechanism does not change the unimodal symmetry property of the probability density function (pdf) of the noisy stochastic gradient. More specifically, the resultant distribution of the effective noise is the convolution of the pdf of symmetric unimodal sub-sampling noise and the Gaussian perturbation noise for privacy.*

We next specialize the convergence result by invoking the amount of local perturbation for privacy according to the Gaussian mechanism in the following corollary.

Corollary 2. *(Utility guarantee) Suppose the global loss function $F(\mathbf{w})$ satisfies the above assumptions. Then, for some constant $c > 0$, a number of iterations T, a batch size $n_b = T/c$, and a learning rate $\eta = 1/\sqrt{\|\mathbf{L}\|_1 n_b}$, the* *convergence rate of the private wireless FL algorithm via the Gaussian mechanism* (3) *is*

$$\frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\|\mathbf{g}^{(t)}\|_1\right] \le \sqrt{\frac{\|\mathbf{L}\|_1}{Tc}}\left(1+\frac{2}{K\cdot\mathrm{SNR}}\right)(R+c/2)$$

$$+\frac{\sqrt{8}c}{3\sqrt{T}}\left(\|\sigma_g\|_1+\frac{2dC}{\epsilon}\sqrt{2\log\frac{1.25}{\delta}}\right). \qquad (7)$$

Remark 3. *It is worth mentioning that we perturb each coordinate of the stochastic gradient via the same amount of perturbation noise. As shown in Theorem 2, the convergence bound gives a fine-grained structure on the error. One may utilize this by perturbing each coordinate independently but differently to obtain the same level of privacy. However, this requires a new analysis of the Gaussian mechanism that accounts for the per coordinate sensitivity.*

IV. Numerical Results

In this section, we complement our theoretical findings through numerical experiments. We consider an image classification task with $K = \{20, 50\}$ users. For the channel model, we use the ITU Extended Pedestrian A model [22]. To capture the long-term variations, we independently regenerate the channels between the PS and the users for each communication round. The subcarrier spacing, the sample rate, and the inverse DFT (IDFT) size are 15 kHz, 30.72 Msps, and 2048, respectively. For encoding, we use 1200 subcarriers (i.e., the signal bandwidth is 18 MHz) and set the rest of subcarriers as guards. Therefore, each OFDM symbol encodes 600 gradients with the proposed scheme. For the time synchronization errors, we assume that the maximum time difference between arriving users' signals is 55.6 ns, and the synchronization uncertainty at the PS is 3 samples, i.e., 97.6 ns. We set $1/\sigma_m^2$ to 20 dB.

For the users' local data, we use the MNIST dataset containing labeled handwritten-digit images size of 28×28 from digit 0 to digit 9. To prepare the data, we first choose 50000 training images from the database, where each digit has 5000 distinct images, and we assume that each user has 250 and 100 distinct images for each digit for $K = 20$ and $K = 50$, respectively. We use 10000 test samples available in the MNIST dataset for the test accuracy calculations. For the model, we consider a convolution neural network (CNN), which has 123090 learnable parameters, given in [10, TABLE I]. We set the learning rate and the batch size to be 0.001 and 64, respectively.

In Fig. 3 and 4, we show the impact of local perturbation noise on the testing accuracy while taking into account the time-synchronization errors (see also Fig. 5 and 6 for training losses comparison). First, we observe that the accuracy performance of the two cases looks similar for two reasons: (i) the stochasticity-induced local perturbation noise is dominant, which is also designed to be the same across users, and (ii) the negligence of large-scale fading[1] for the wireless channel. However, the privacy guarantees for $K = 50$ users are improved, as shown in (5). This is because the local perturbation noises get aggregated over the wireless channel. As a numerical example, by invoking worst case expression for the privacy leakage in (5) when $\sigma^2 = 0.1$, yields that the achieved local privacy level for $K = 20$ users is $\epsilon = 9.3262$ and for $K = 50$, $\epsilon = 6.5936$. Although the obtained privacy guarantees seem loose under the choice of simulation parameters, one can utilize the random client participation in federated learning to tighten further and amplify the privacy guarantees. Additionally, in a typical wireless FL setting, we may have hundreds of mobile users; the PS can select a specified number of users at each communication round. Selecting K out of N users uniformly at random will amplify the privacy guarantee as $(\log(p(e^\epsilon - 1) + 1), p\delta)$, where $p = K/N$. We also compare the performance of our proposed FSK-MV scheme with the OBDA scheme [7]. We can see clearly that the imperfect time-synchronization cause a drastic reduction in the performance of the OBDA scheme. On the other hand, the FSK-MV scheme is robust against time-synchronization errors.

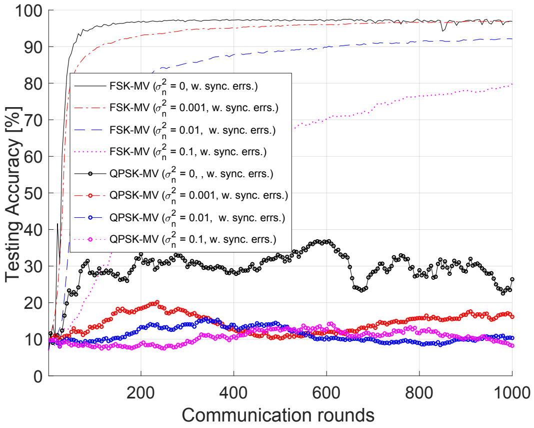

Fig. 3. The impact of local perturbation noise on the testing accuracy, where $\delta = 10^{-3}$, $K = 20$ users, $C = 1$ and $T = 1000$ communication rounds.

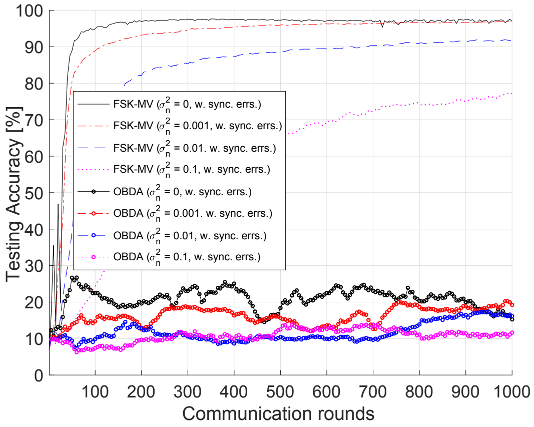

Fig. 4. The impact of local perturbation noise on the testing accuracy, where $\delta = 10^{-3}$, $K = 50$ users, $C = 1$ and $T = 1000$ communication rounds.

[1] As future work, we will study the impact of large scale fading and take the transmit power control into account.

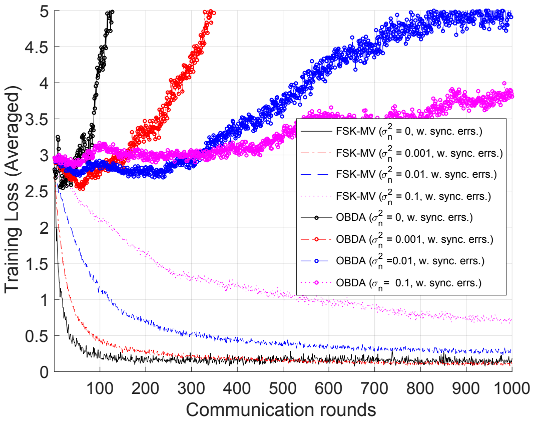

Fig. 5. The impact of local perturbation noise on the training loss, where $\delta = 10^{-3}$, $K = 20$ users, $C = 1$ and $T = 1000$ communication rounds.

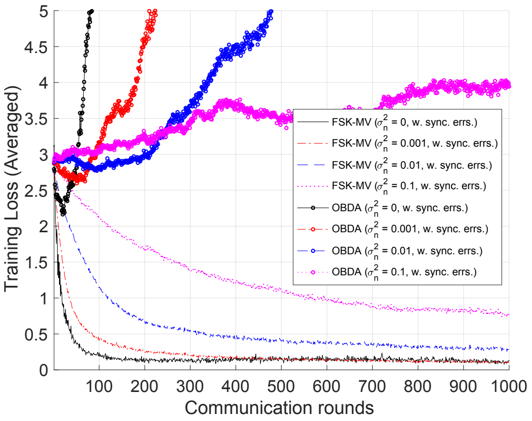

Fig. 6. The impact of local perturbation noise on the training loss, where $\delta = 10^{-3}$, $K = 50$ users, $C = 1$ and $T = 1000$ communication rounds.

V. Conclusion & Future Work

In this work, we have studied the problem of wireless federated learning with local differential privacy constraints. Further, we have proposed a non-coherent scheme based on FSK-MV modulation that is robust to time synchronization errors and does not require channel state information or power control. We have formally characterized the per-user privacy leakage and shown that our proposed scheme boosted the privacy guarantees, and further that that the leakage scales as $\mathcal{O}(1/\sqrt{K})$ thanks to the superposition property of the wireless channel. Furthermore, we have acknowledged the ongoing efforts in the field to enhance convergence by employing model sparsification. In our ongoing work, we plan to consider this technique as a potential approach to improving performance. Finally, we mention the following open question: if we relax the *global* CSI assumption in the privacy analysis at the untrusted node (i.e., the PS), we get an improved level of privacy compared to the privacy guarantee presented in (4)?

References

[1] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 491–506, 2019.

[2] K. Yang, T. Jiang, Y. Shi, and Z. Ding, "Federated learning via over-the-air computation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 3, pp. 2022–2035, 2020.

[3] M. S. E. Mohamed, W.-T. Chang, and R. Tandon, "Privacy amplification for federated learning via user sampling and wireless aggregation," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 39, no. 12, pp. 3821–3835, 2021.

[4] D. Liu and O. Simeone, "Privacy for free: Wireless federated learning via uncoded transmission with adaptive power control," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 39, no. 1, pp. 170–185, 2020.

[5] Y. Shao, D. Gündüz, and S. C. Liew, "Federated edge learning with misaligned over-the-air computation," *arXiv preprint arXiv:2102.13604*, 2021.

[6] A. Şahin and R. Yang, "A survey on over-the-air computation," *IEEE Communication Surveys and Tutorials*, pp. 1–33, 2023.

[7] G. Zhu, Y. Du, D. Gündüz, and K. Huang, "One-bit over-the-air aggregation for communication-efficient federated edge learning: Design and convergence analysis," *IEEE Transactions on Wireless Communications*, vol. 20, no. 3, pp. 2120–2135, 2020.

[8] M. M. Amiri, T. M. Duman, D. Gündüz, S. R. Kulkarni, and H. V. Poor, "Blind federated edge learning," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5129–5143, 2021.

[9] M. Goldenbaum and S. Stanczak, "Robust analog function computation via wireless multiple-access channels," *IEEE Transactions on Communications*, vol. 61, no. 9, pp. 3863–3877, 2013.

[10] A. Şahin, "Distributed learning over a wireless network with non-coherent majority vote computation," *IEEE Transactions on Wireless Communications*, pp. 1–16, 2023.

[11] S. Hoque, M. H. Adeli, and A. Şahin, "Chirp-based over-the-air computation for long-range federated edge learning," in *Proceedings of the 2022 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2022, pp. 1–7.

[12] J. Bernstein, Y.-X. Wang, K. Azizzadenesheli, and A. Anandkumar, "signSGD: Compressed optimisation for non-convex problems," in *Proceedgins of the International Conference on Machine Learning (ICML)*. PMLR, 2018, pp. 560–569.

[13] A. Şahin, "A demonstration of over-the-air-computation for FEEL," in *Proceedings of IEEE Global Communications Conference Workshops (GLOBECOM Workshop) - Edge Learning over 5G Mobile Networks and Beyond*, Dec. 2022, pp. 1–7.

[14] C. Dwork, A. Roth *et al.*, "The algorithmic foundations of differential privacy," *Foundations and Trends® in Theoretical Computer Science*, vol. 9, no. 3–4, pp. 211–407, 2014.

[15] M. Joseph, A. Roth, J. Ullman, and B. Waggoner, "Local differential privacy for evolving data," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2018, pp. 2375–2384.

[16] R. C. Geyer, T. Klein, and M. Nabi, "Differentially private federated learning: A client level perspective," *arXiv preprint arXiv:1712.07557*, 2017.

[17] O. Choudhury, A. Gkoulalas-Divanis, T. Salonidis, I. Sylla, Y. Park, G. Hsu, and A. Das, "Differential privacy-enabled federated learning for sensitive health data," *arXiv preprint arXiv:1910.02578*, 2019.

[18] M. Seif, R. Tandon, and M. Li, "Wireless federated learning with local differential privacy," in *Proceedings of the 2020 IEEE International Symposium on Information Theory (ISIT)*, 2020, pp. 2604–2609.

[19] B. Hasırcıoğlu and D. Gündüz, "Private wireless federated learning with anonymous over-the-air computation," in *Proceedings of the 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021, pp. 5195–5199.

[20] Y. Yang, Y. Zhou, Y. Wu, and Y. Shi, "Differentially private federated learning via reconfigurable intelligent surface," *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 19 728–19 743, 2022.

[21] J. J. Bussgang, "Crosscorrelation functions of amplitude-distorted Gaussian signals," *Research Laboratory of Electronics, Massachusetts Institute of Technology*, 1952.

[22] ETSI, "Relay radio transmission and reception (TS 136 116 V17.0.0)," *Technical Report*, Apr. 2022.