Communication-Efficient Federated Learning Over-the-Air with Sparse One-Bit Quantization

Junsuk Oh, Donghyun Lee, Dongwook Won, Wonjong Noh, and Sungrae Cho

***Abstract*—Federated learning (FL) is a framework for realizing distributed machine learning in an environment where training samples are distributed to each device. Recently, FL has employed over-the-air computation enabling all devices to transmit learning model updates simultaneously. This work proposes a communication-efficient sparse one-bit analog aggregation (SOBAA) method, incorporating new power control, layer-wise scaled one-bit quantization, layer-wise sparsification, and an error-feedback mechanism. We derive a tight upper bound of the expected convergence rate of the proposed SOBAA as a closed-form expression. From this expression, we explicitly identify the relationship between the convergence rate and compression and aggregation errors. Based on the theoretical convergence analysis, we formulate a joint optimization problem of the compression ratio and power control to minimize compression and aggregation errors, leading to the fastest convergence. In each communication round, the optimization problem is decomposed, and solved in a computationally efficient and feasible way. From this solution, we characterize the trade-off between learning performance and communication cost. Through extensive experiments on well-known MNIST and CIFAR-10 datasets, we confirm that the proposed method provides an enhanced trade-off performance between test accuracy and communication costs and a faster convergence rate than the other state-of-the-art methods. In addition, it is proven that the proposed method is more effective for more complex datasets and learning models.**

***Index Terms*—Analog aggregation, federated learning over-the-air, power control, sparse one-bit quantization**

I. Introduction

MACHINE learning, which has proliferated in recent years, provides various application services through big data. However, centralized data center learning faces privacy challenges because a parameter server (PS) collects data generated on each device [1]. Federated learning (FL) is a machine learning paradigm that has emerged to address these challenges and aims to train one global learning model using data stored on numerous local devices based on adjustments of the PS [2]. FL repeats a communication round in which local gradients of trained local learning models are aggregated until the global learning model converges. Thus, raw data are not collected therefore, FL offers an opportunity to leverage the computational resources of each device while protecting privacy, unlike centralized data center learning.

In FL, all types of learning models can be selected depending on the provided application services. However, the numerous parameters that comprise the learning model can lead to excessive communication costs that cause bottlenecks [3]. For example, artificial neural networks, such as ResNet-152 [4] with 60.3M parameters and VGG-16 [5] with 138M parameters, must be selected to provide image classification services. Thus, gradient compression schemes have been considered in FL to reduce communication costs per device.

First, quantization is a scheme that limits each parameter to be represented by a few finite bits in the local gradient. Examples of such schemes include eight-bit quantization [6], quantized stochastic gradient descent (QSGD) [7], and signSGD [8]. Second, sparsification is a scheme that removes some parameters and retains only critical parameters in the local gradient. Examples of such schemes include top-k [9], sketched-SGD [10], model pruning [11], and dropout [12]. Third, a hybrid scheme combines quantization with sparsification. Examples of such schemes include hard-threshold SGD [13], sketchML [14], and Qsparse-local-SGD [15]. Moreover, various gradient compression schemes include low rank [16], communication censoring [17], and compressive sensing [18].

However, information loss due to local gradient compression leads to errors in global gradient estimation, degrading the convergence rate and learning performance. Thus, the gradient compression schemes tend to incorporate an error-feedback mechanism [19] that compensates for this information loss in the following communication round [20], [21].

In contrast, despite the gradient compression scheme, FL still faces bottlenecks because communication cost is also affected by the number of devices participating in each communication round. Thus, FL has recently evolved into FL over the air (FLOA) [22], [23] by integrating it with over-the-air computation (OAC) [24], which takes advantage of the waveform superposition property of the multiple access channel (MAC) to convert the air to computers that perform computing and communication functions. In FLOA, over the same time-frequency resources, all devices simultaneously transmit local gradients, which are analog aggregated in the air, and the PS receives the aggregated local gradients. Thus, it enables FLOA to achieve higher bandwidth efficiency than conventional FL by integrating communication and computation. However, FLOA has some representative challenges: high communication costs caused by numerous learning model parameters, and signal distortion caused by noise and fading.

This work was supported in part by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2024-RS-2022-00156353) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation) and in part by the National Research Foundation of Korea (NRF) grants funded by the Korean government (MSIT) (RS-2023-00209125, NRF-2023R1A2C1003003). *(Corresponding authors: Wonjong Noh; Sungrae Cho.)*

Junsuk Oh, Donghyun Lee, Dongwook Won, and Sungrae Cho are with the School of Computer Science and Engineering, Chung-Ang University, Seoul 06974, South Korea (e-mail: jsoh@uclab.re.kr; dhlee@uclab.re.kr; dwwon@uclab.re.kr; srcho@cau.ac.kr).

Wonjong Noh is with the School of Software, Hallym University, Chuncheon 24252, South Korea (e-mail: wonjong.noh@hallym.ac.kr).

To address these problems, in this work, we propose an FLOA-based new compression and transmission approach.

A. Related Work

In FLOA, high communication costs and signal distortion are critical problems. Therefore, numerous prior studies have focused on following types of research: gradient compression [25]–[31] and power control [22], [23], [32]–[37].

First, some works have considered gradient compression schemes to mitigate high communication costs. The authors of [25] treated each iteration of distributed SGD (DSGD) as a distributed over-the-air lossy computation problem and remarked that the transmission of learning model parameters in [22] and [23] results in waste of power. Thus, considering the physical layer resource constraints in narrow-band channels, they proposed analog DSGD (A-DSGD). This source-coding algorithm uses an error-feedback mechanism, gradient sparsification, and random linear projection for gradient dimension reduction. Furthermore, to extend the work of [25] considering additive white Gaussian noise (AWGN) MAC to fading MAC, they proposed compressed A-DSGD (CA-DSGD) in [26]. However, the unencoded analog modulation used for over-the-air aggregation in [25] and [26] is challenging to deploy in digital modulation-based systems. Therefore, to be compatible with these systems, the authors of [27] proposed one-bit broadband digital aggregation (OBDA) based on a one-bit quantization, digital modulation, and the majority vote of the PS. Moreover, they analyzed the convergence rate by the effect of the wireless channel hostilities. The authors of [28] remarked on the degradation of the convergence rate and learning performance via one-bit quantization in [27] and proposed error-feedback OBDA (EFOBDA). Inspired by the works in [25], [26], and [27], the authors of [29] proposed a one-bit compressive sensing analog aggregation (OBCSAA). Furthermore, they provided the relationship between FL and analog aggregation with this method via convergence analysis. Moreover, noting that random linear projection in the work of [25] and [26] increase the computational complexity of the receiver, the authors of [30] proposed a time-correlated sparsification with hybrid aggregation (TCS-H) method, in which the global sparse gradient (which same between devices) and the local sparse gradient (which may differ between devices) are aggregated through OAC and digital communication, respectively. Inspired by the works in [25] and [26], the authors of [31] proposed the over-the-air federated multi-task learning (OA-FMTL) by designing a transmission method called model sparsification and random compression and a reception method called modified turbo compressive sensing.

Second, other works have proposed power control design to mitigate signal distortion.[1] Some works have considered device scheduling strategy in power control design to exclude some devices that cannot mitigate signal distortion depending on the channel state within limited transmission power. For example, the authors of [22] designed a broadband analog aggregation method by deriving a trade-off between learning and communication for power control and device scheduling. Specifically, the authors introduced a truncated channel inversion approach to exclude devices and some local gradients according to their channel states. In another study [23], the authors proposed a joint design for receiver beamforming and device scheduling to maximize the number of participating devices by minimizing the requirements of a given mean squared error in a multi-antenna scenario. Considering energy-constrained devices, the authors of [32] proposed an energy-aware device scheduling strategy to maximize the average number of participating devices. Furthermore, the relationship between FL and analog aggregation is unclear therefore, another study [33] quantified the effect of OAC on FL and proposed the joint optimization of power control and device scheduling. Considering analog downlink and uplink communications, the authors of [34] proposed a semidefinite programming technique-based joint optimization scheme of power control and device scheduling along with an adaptive reweighing scheme. The remaining works concentrating on the power control design without a device scheduling strategy are as follows. The authors of [35] proposed a gradient-aware power control design to minimize individual signal distortions for wireless transmission. The authors of [36] characterized the effects of the aggregation errors for the bias and mean squared error and proposed an optimal power control design. Similarly, the authors of [37] theoretically derived an upper bound of the optimality gap in the case of aggregation errors and proposed optimal transmission power control to minimize the optimality gap under individual power constraints.

[1] In the analog aggregation model of the FLOA, mitigating signal distortion caused by noise and fading cannot be ignored in all FLOA-related works [38]; therefore, we remark that all works related to gradient compression considered channel inversion-based power control.

B. Motivation, Contribution, and Organization

In FLOA, gradient compression, and power and error control play a very important role in performance. However, prior studies have adopted a fixed compression rate, and many joint optimal controls are performed based on it, which motivated this study. To the best of our knowledge, this is the first study that introduces a dynamic control of the sparsification-based compression ratio and performs joint optimal power and error control with it in FLOA. Table I summarizes the notable differences between the proposed and existing methods. The main contributions of this paper are outlined as follows:

- For the communication-efficient FLOA, we propose the sparse one-bit analog aggregation (SOBAA) method. First, the SOBAA employs a new synchronized layer-wise sparsification, which drives all devices to sparsify the same layers simultaneously. In FLOA, the gradient compression comes from gradient quantization, not from sparsification. However, if the proposed layer-wise sparsification is employed, it can contribute to gradient compression. Furthermore, the gradient compression ratio by layer-wise sparsification can be dynamically controlled or optimized according to the channel status. Second, inspired by [39], the SOBAA develops layer-wise scaled one-bit quantization, advantaging that it generally has an improved convergence rate due to lower compression

TABLE I: FLOA considering gradient compression and power control

Method	Gradient compression	Power control	Evaluation metrics
A-DSGD [25]	Top-k, random projection, compressive sensing, error-feedback mechanism	Channel inversion (AWGN)	Test accuracy
CA-DSGD [26]	Top-k, random projection, compressive sensing, error-feedback mechanism	Truncated channel inversion (fading with perfect and imperfect CSI)	Test accuracy
OBDA [27]	One-bit quantization	Truncated channel inversion (AWGN, fading with perfect and imperfect CSI)	Train loss, test accuracy, communication latency
EFOBDA [28]	One-bit quantization, error-feedback mechanism	Channel inversion, convex optimization method (AWGN, fading with perfect CSI)	Train loss, test accuracy
OBCSAA [29]	One-bit quantization, top-k, random projection, compressive sensing	Channel inversion, convex optimization method (fading with perfect CSI)	Train loss, test accuracy, communication cost
TCS-H [30]	Top-k, error-feedback mechanism	Channel inversion (fading with perfect CSI)	Test accuracy, communication cost
OA-FMTL [31]	Top-k, random projection, compressive sensing, error-feedback mechanism	Channel inversion, convex optimization method (fading with perfect CSI)	Test accuracy, running time, communication cost
Proposed	Layer-wise scaled one-bit quantization, layer-wise sparsification, error-feedback mechanism	Channel inversion, convex optimization method (fading with perfect CSI)	Train loss, test accuracy, communication cost

error than entirely scaled or unscaled one-bit quantization schemes [19], [40]. It optimizes the layer-wise sparsification using the estimated layer-wise magnitude scaling factors. Third, the SOBAA employs analog aggregation over the air for information transmission. In particular, the SOBAA indirectly carries the information on magnitude scaling factors by controlling the power amplitude of the signal, instead of using additional bits.

- To mitigate the compression error, we consider the error-feedback mechanism in the proposed SOBAA. Thus, to show the effectiveness of the error-feedback mechanism, we derive a tight upper bound on the expected convergence rate of the proposed SOBAA with this mechanism (SOBAA-EFO) and without this mechanism (SOBAA-EFX) as a closed-form expression. Through this expression, we explicitly identify the relationship between convergence rate and compression and aggregation errors.
- Based on this theoretical convergence analysis, we formulate a joint optimization problem that determines the optimal compression rate and controls the transmit power to minimize compression and aggregation errors, leading to the fastest convergence. In each communication round, the optimization problem is decomposed and solved in a computationally efficient and feasible way. From this solution, we characterize the trade-off between learning performance and communication costs.
- We perform extensive experimental evaluations using well-known MNIST and CIFAR-10 datasets and confirm that, compared with the state-of-the-art approaches, the proposed approach provides a faster convergence rate and an enhanced trade-off in terms of communication cost and test accuracy. The proposed approach is also proven to be robust for the variations of device populations or network conditions. In particular, it works more efficiently for complex datasets and local learning models that comprise more layers and parameters.

The remainder of this paper is organized as follows. First, Section II proposes the system model of the proposed SOBAA method. Section III derives the expected convergence rate as a closed-form expression for convergence analysis. Next, Section IV formulates a joint optimization problem to determine the compression ratio and transmission power and introduces the corresponding solutions to minimize communication costs and compression and aggregation errors. Finally, the experimental results and their setups are provided in Section V, followed by a conclusion and discussion in Section VI.

II. Proposed FLOA System Model

For wireless networks, we consider an FL system consisting of a single PS and K devices. Thus, the PS and all devices aim to train one global learning model collaboratively by exploiting gradient compression and analog aggregation, as presented in Fig. 1.

A. Learning Model

In the FL system, device k has its local dataset consisting of D_k data samples. For training the learning model, the local loss function $\mathcal{F}_k(\cdot)$ with parameter vector $\mathbf{w}$ is defined as

$$\mathcal{F}_k(\mathbf{w}) = \frac{1}{D_k}\sum_{d=1}^{D_k}\mathcal{F}_{k,d}(\mathbf{w};\mathbf{x}_{k,d},y_{k,d}), \tag{1}$$

where $\mathcal{F}_{k,d}(\cdot)$ denotes the loss function on sample d consisting of feature vector $\mathbf{x}_{k,d}$ and label $y_{k,d}$. Then, based on $\mathcal{F}_k(\cdot)$ for all devices, the global loss function $\mathcal{F}(\cdot)$ with $\mathbf{w}$ is defined as

$$\mathcal{F}(\mathbf{w}) = \frac{1}{D}\sum_{k=1}^{K} D_k\mathcal{F}_k(\mathbf{w}), \tag{2}$$

where $D = \sum_{k=1}^{K} D_k$. Thus, the FL system aims to determine an optimal parameter vector $\mathbf{w}^*$ by minimizing $\mathcal{F}(\cdot)$ as

$$\textbf{P1:}\quad \mathbf{w}^* = \arg\min_{\mathbf{w}} \quad \mathcal{F}(\mathbf{w}). \tag{3}$$

The FL system adjusts the training process between the PS and devices to be configured in a distributed manner, avoiding collecting raw samples directly from each device to the PS. In each communication round $t \in \{1,\ldots,T\}$, the PS broadcasts parameter vector $\mathbf{w}^t$ to all devices. Each device

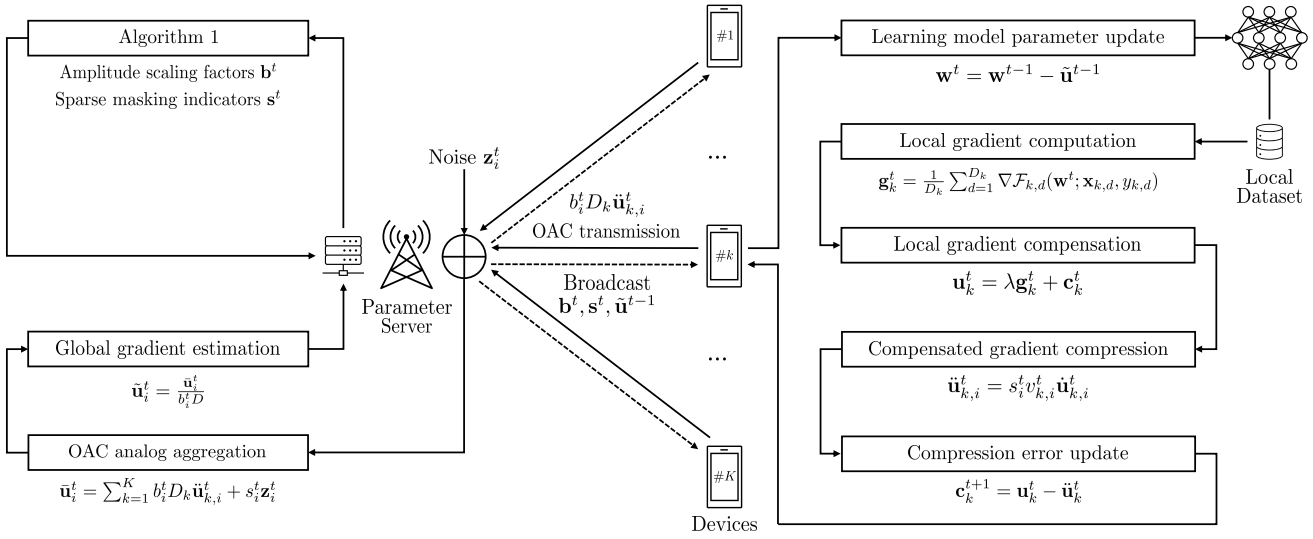


Fig. 1: FLOA system model with SOBAA.

updates the received $\mathbf{w}^t$ based on its local dataset using a first-order approximation algorithm to minimize the local loss function and derives local gradient vector $\mathbf{g}_k^t$ as

$$\mathbf{g}_k^t = \frac{1}{D_k} \sum_{d=1}^{D_k} \nabla \mathcal{F}_{k,d}(\mathbf{w}^t; \mathbf{x}_{k,d}, y_{k,d}), \tag{4}$$

where $\nabla \mathcal{F}_{k,d}(\cdot)$ denotes the gradient of $\mathcal{F}_{k,d}(\cdot)$ with respect to $\mathbf{w}^t$. Then, each device transmits $\mathbf{g}_k^t$, and the PS aggregates that to derive the global gradient vector $\mathbf{g}^t$ as

$$\mathbf{g}^t = \frac{1}{D} \sum_{k=1}^{K} D_k \mathbf{g}_k^t. \tag{5}$$

Therefore, the PS updates $\mathbf{w}^t$ as

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \lambda \mathbf{g}^t, \tag{6}$$

where λ denotes the learning rate. The FL system iterates the training process including (4), (5), and (6) until the convergence condition is satisfied or the maximum communication round T is reached.

B. Compression Model

For the communication-efficient FL system, we introduced a layer-wise scaled one-bit quantization and layer-wise sparsification. However, the aggressive gradient compression degrades the convergence rate and learning performance. Thus, we incorporated the error-feedback mechanism to mitigate the negative effect of the compression error. To implement this in the FL system, we partitioned the local gradient vector into blocks (i.e., subvectors) and derived the (gradient) magnitude scaling factor block-wise, which is the average of the parameters in the block. However, if a block comprises parameters of different layers, the magnitude scaling factor of that block may not correctly indicate its importance because the magnitude of the parameters between layers can be significantly different [41]. To this end, we considered layers to be blocks.

In each communication round t, each device derives the compensated local gradient vector $\mathbf{u}_k^t$ by adding a compression error vector $\mathbf{c}_k^t$ as

$$\mathbf{u}_k^t = \lambda \mathbf{g}_k^t + \mathbf{c}_k^t, \tag{7}$$

where $\mathbf{c}_k^t$ compensates for information loss caused by gradient compression in the previous round. Then, each device derives the magnitude scaling factor $v_{k,i}^t$ layer-wise as

$$v_{k,i}^t = \frac{1}{J_i} \left\| \mathbf{u}_{k,i}^t \right\|_1, \tag{8}$$

where $i \in \{1, \ldots, I\}$ denotes the layer i. In addition, each device derives the quantized $\mathbf{u}_k^t$ parameter-wise as

$$\dot{u}_{k,i,j}^t = \begin{cases} +1, & u_{k,i,j}^t \geq 0, \\ -1, & \text{otherwise}, \end{cases} \tag{9}$$

where $j \in \{1, \ldots, J_i\}$ represents the parameter j of the layer i. Thus, the compressed $\mathbf{u}_k^t$ is defined as

$$\ddot{\mathbf{u}}_{k,i}^t = s_i^t v_{k,i}^t \dot{\mathbf{u}}_{k,i}^t, \tag{10}$$

where s_i^t denotes the sparse masking indicator (i.e., $s_i^t = 0$ if the layer i is not transmitted at each round t, and $s_i^t = 1$, otherwise). Thus, $\mathbf{c}_k^t$ is updated as

$$\mathbf{c}_k^{t+1} = \mathbf{u}_k^t - \ddot{\mathbf{u}}_k^t, \tag{11}$$

where $\mathbf{c}_k^{t+1}$ is locally stored on device k.

Remark 1. From [42], it is known that aggregating layers with small gradients cannot improve the test accuracy significantly and only increases communication costs, i.e., transmitting only some layers with higher gradients lowers costs while maintaining test accuracy compared to transmitting all layers.

Remark 2. The proposed layer-wise sparsification operates in a synchronized manner, that is, all devices' same specific layers are sparsified for each round. Through this, the averaged gradient information at PS can be transmitted without losing sparsity to the downlink. It can reduce the amount of downlink traffic and enhance the drop-out effect in terms of information abstraction by each layer in the learning model.

C. Aggregation Model

Based on the observation that the PS requires only aggregated local gradients, FL evolved to FLOA by incorporating OAC, exploiting the waveform superposition property of the MAC. In the FLOA system, analog aggregation enables the simultaneous transmission of local gradients for all devices to the PS over the same uplink time-frequency channels.

In each communication round t, each device transmits $\ddot{\mathbf{u}}_{k,i}^t$, which is supposed to be aggregated to derive $\ddot{\mathbf{u}}_i^t$ layer-wise as

$$\ddot{\mathbf{u}}_i^t = \frac{1}{D}\sum_{k=1}^{K} D_k \ddot{\mathbf{u}}_{k,i}^t. \tag{12}$$

However, since $\ddot{\mathbf{u}}_{k,i}^t$ is transmitted in an analog manner and aggregated over the air, the received signal subvector $\bar{\mathbf{u}}_i^t$ on the PS is given by

$$\bar{\mathbf{u}}_i^t = \sum_{k=1}^{K} h_k^t p_{k,i}^t \ddot{\mathbf{u}}_{k,i}^t + s_i^t \mathbf{z}_i^t, \tag{13}$$

where h_k^t denotes the channel coefficient between the device k and the PS, modeled as i.i.d. Rayleigh fading by generating from $\mathcal{N}(0,1)$.[2] In addition, $p_{k,i}^t$ represents the transmit power, and $\mathbf{z}_i^t$ denotes the AWGN subvector (i.e., $z_{i,j}^t \sim \mathcal{N}(0,\sigma^2)$, where σ^2 indicates the noise power).

In order to ensure that (13) can be approximate to (12) as closely as possible by mitigating the signal distortion caused by the noise and fading, we designed $p_{k,i}^t$ as

$$p_{k,i}^t = \frac{b_i^t D_k}{h_k^t}, \tag{14}$$

where b_i^t is the (power) amplitude scaling factor. The transmit power constraint must be satisfied on each device as

$$\left|p_{k,i}^t \ddot{u}_{k,i,j}^t\right|^2 = \left|\frac{b_i^t D_k s_i^t v_{k,i}^t}{h_k^t}\right|^2 \le P, \tag{15}$$

where P denotes the maximum transmit power, and $\dot{u}_{k,i,j}^t$ is eliminated because of $\dot{u}_{k,i,j}^t = \pm 1$.

By substituting (14) into (13), $\bar{\mathbf{u}}_i^t$ is rewritten as

$$\bar{\mathbf{u}}_i^t = \sum_{k=1}^{K} b_i^t D_k \ddot{\mathbf{u}}_{k,i}^t + s_i^t \mathbf{z}_i^t, \tag{16}$$

where D_k represents the pre-processing factor. Upon $\bar{\mathbf{u}}_i^t$, the PS derives the estimated $\ddot{\mathbf{u}}_i^t$ as

$$\tilde{\mathbf{u}}_i^t = \frac{\bar{\mathbf{u}}_i^t}{b_i^t D} = \ddot{\mathbf{u}}_i^t + \frac{s_i^t \mathbf{z}_i^t}{b_i^t D}, \tag{17}$$

where $b_i^t D$ denotes the post-processing factor. Then, the PS broadcasts $\tilde{\mathbf{u}}^t$ to all devices to update parameter vector as

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \tilde{\mathbf{u}}^t, \tag{18}$$

where we assumed that all devices receive perfect $\tilde{\mathbf{u}}^t$ through error-free downlink channels by the high transmit power and bandwidth of the PS.

[2]For the sake of simplicity, we assumed the perfect uplink channel state information (CSI) and that the CSI of the block fading channels is unchanged in each round t, but may be changed independently over different rounds.

III. Convergence Analysis

This section formally characterizes the effects of layer-wise scaled one-bit quantization, layer-wise sparsification, the error-feedback mechanism, and an analog aggregation on the FLOA system by analyzing its convergence rate.

A. Basic Assumptions

To facilitate convergence analysis, we made widely adopted assumptions regarding the compression operator, loss function, and gradient vector.

Assumption 1. *(Compression Operator):* The operator $\mathcal{C}(\cdot): \mathbb{R}^J \to \mathbb{R}^J$ denotes a δ-approximate compression operator with approximation factor $\delta \in [0,1]$, that is,

$$\|\mathcal{C}(\mathbf{u}) - \mathbf{u}\|_2^2 \le (1-\delta)\|\mathbf{u}\|_2^2, \quad \forall \mathbf{u} \in \mathbb{R}^J, \tag{19}$$

where $J = \sum_{i=1}^{I} J_i$. In addition, this is extended and defined in the layer-wise form as follows:

$$\begin{aligned}\|\mathcal{C}(\mathbf{u}_i) - \mathbf{u}_i\|_2^2 &= \|s_i v_i \dot{\mathbf{u}}_i - \mathbf{u}_i\|_2^2 \\ &= J_i s_i (v_i)^2 - 2 s_i v_i \|\mathbf{u}_i\|_1 + \|\mathbf{u}_i\|_2^2 \\ &= \left(1 - \frac{J_i s_i (v_i)^2}{\|\mathbf{u}_i\|_2^2}\right)\|\mathbf{u}_i\|_2^2 \\ &\le (1 - s_i\delta)\|\mathbf{u}_i\|_2^2, \quad \forall i. \end{aligned} \tag{20}$$

Assumption 2. *(Loss Function Bound):* The global loss function $\mathcal{F}(\cdot)$ is lower bounded by an optimal parameter vector $\mathbf{w}^*$ for the given parameter vector $\mathbf{w}$, that is,

$$\mathcal{F}(\mathbf{w}) \ge \mathcal{F}(\mathbf{w}^*), \quad \forall \mathbf{w} \in \mathbb{R}^J. \tag{21}$$

Assumption 3. *(Smoothness and Lipschitz Continuity):* The global loss function $\mathcal{F}(\cdot)$ is L-smooth and the corresponding gradient vector $\nabla\mathcal{F}(\cdot)$ is L-Lipschitz continuous, that is,

$$\begin{aligned}&\left|\mathcal{F}(\mathbf{w}') - \left(\mathcal{F}(\mathbf{w}) + \nabla\mathcal{F}(\mathbf{w})^\top(\mathbf{w}' - \mathbf{w})\right)\right| \\ &\le \frac{L}{2}\|\mathbf{w}' - \mathbf{w}\|_2^2, \quad \forall \mathbf{w}', \mathbf{w} \in \mathbb{R}^J, \end{aligned} \tag{22}$$

where L is a non-negative constant. In addition, this is defined in an alternate form as follows:

$$\|\nabla\mathcal{F}(\mathbf{w}') - \nabla\mathcal{F}(\mathbf{w})\|_2 \le L\|\mathbf{w}' - \mathbf{w}\|_2. \tag{23}$$

Assumption 4. *(Gradient Vector Bound):* The local gradient vector $\mathbf{g}_k$ is independent and unbiased estimate of the global gradient $\mathbb{E}[\mathbf{g}] = \nabla\mathcal{F}(\mathbf{w})$, that is,

$$\|\mathbf{g}_k\|_2^2 \le G^2, \quad \forall k, \tag{24}$$

where G is a positive constant. In addition, this is extended and defined in the layer-wise form as follows:

$$\|\mathbf{g}_{k,i}\|_2^2 \le G_i^2, \quad \forall i. \tag{25}$$

B. Main Results

Based on the assumptions, we provided the expected convergence rate of the proposed SOBAA as a closed-form expression. First, we derived **Lemma 1** to describe the compression error caused by layer-wise scaled one-bit quantization, layer-wise sparsification, and the error-feedback mechanism.

Lemma 1. In the FL system, under **Assumptions 1** and **4**, the compression error is bounded by

$$\mathbb{E}\left\|\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^{t+1}\right\|_2^2 \leq \sum_{i=1}^{I}\frac{(1-s_i^t\delta)(2-\delta)\lambda^2G_i^2}{\delta} \times\left((t-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right), \quad (26)$$

where S_i is most recent communication round t with $s_i^t=1$. In addition, $M_i\geq 1$ is the maximum non-transmit period (i.e., $t-S_i\leq M_i$).

Proof: See Appendix A. ∎

In addition, we derived **Lemma 2** to describe the aggregation error caused by analog aggregation.

Lemma 2. In the FLOA system, according to (17) and the variance expression by expected value, the aggregation error is given by

$$\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2=\sum_{i=1}^{I}\mathbb{E}\left\|\frac{s_i^t\mathbf{z}_i^t}{b_i^tD}\right\|_2^2=\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2}{(b_i^t)^2D^2}. \quad (27)$$

∎

Based on **Lemmas 1** and **2**, we represented the convergence rate of the proposed SOBAA-EFO in **Theorem 1**.

Theorem 1. Under **Assumptions 2**-**4**, the convergence rate of SOBAA-EFO is given by

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2 \leq \frac{2(\mathcal{F}(\mathbf{w}^1)-\mathcal{F}(\mathbf{w}^*))}{\delta\lambda T}+\frac{G^2}{\delta} +\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2}{\delta(b_i^t)^2\lambda^2D^2} +\frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{I}\frac{(1-s_i^t\delta)G_i^2}{\delta^2} \times\left((t-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right). \quad (28)$$

Proof: See Appendix B. ∎

To show the effectiveness of the error-feedback mechanism, based on **Lemma 2**, we also represented the convergence rate of SOBAA-EFX in **Theorem 2**.

Theorem 2. Under **Assumptions 1**-**4**, the convergence rate of SOBAA-EFX is given by

$$\frac{1}{T}\sum_{t=1}^{T}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2 \leq \frac{2(\mathcal{F}(\mathbf{w}^1)-\mathcal{F}(\mathbf{w}^*))}{\lambda T} +\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2}{(b_i^t)^2\lambda^2D^2} +\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}(1-s_i^t\delta)G_i^2. \quad (29)$$

Proof: See Appendix D. ∎

According to observations on **Theorems 1** and **2**, the FLOA system can achieve an improved convergence performance by minimizing total error in the upper bound. On the right-hand side (RHS) of (28), the second term is constant, which is the weighted upper bound of the gradient, and other terms decrease as maximum communication round T increases. Notably, the third and last terms denote the weighted aggregation error (WAE) and weighted compression error (WCE), respectively. Similarly, on the RHS of (29), all terms decrease as maximum communication round T increases, and the second term and last term denote WAE and WCE, respectively.

WAE (i.e., the negative effect of the noise) decreases as the amplitude scaling factor b_i^t increases. In addition, the effect is eliminated when the sparse masking indicator $s_i^t=0$ (i.e., the layer i is not transmitted). However, WCE decreases when $s_i^t=1$ (i.e., the layer i is transmitted). That is, when $s_i^t=0$, WAE decreases, and WCE increases. In contrast, when $s_i^t=1$, vice versa. Therefore, s_i^t and b_i^t determine WAE and WCE, affecting the learning performance and communication costs. Using this theoretical convergence analysis, we proposed the joint optimization design that determines the compression ratio via s_i^t and controls the transmit power via b_i^t to minimize the total error in the upper bound of the convergence rate. Using the solution, we characterize the trade-off between learning performance and communication costs.

IV. Joint Minimization of Errors and Costs

For the communication-efficient FLOA system, this section formulates the joint optimization problem and derives computationally feasible solutions.

A. Formulation of Joint Optimization Problem

For SOBAA-EFO, the PS aims to determine all amplitude scaling factors $\mathbf{b}^t=\{b_i^t\}_{i=1}^I$ and all sparse masking indicators $\mathbf{s}^t=\{s_i^t\}_{i=1}^I$ to minimize (28) in each communication round t. Thus, by eliminating the common factors after dropping the irrelevant terms in (28), we defined the total error e^t as

$$e^t=\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2}{(b_i^t)^2\lambda^2D^2}+\sum_{i=1}^{I}\frac{(1-s_i^t\delta)\,G_i^2}{\delta} \times\left((t-S_i)^2+\frac{(2-\delta)\,M_i^2}{\delta}\right), \quad (30)$$

where we can find two observations: *1)* For $\ddot{u}_{i,j}^t$ of (17) not to be significantly distorted by the noise, based on the variance

expression by expected value, the post-processed noise needs to be reduced as

$$\mathbb{E}\left|\frac{z_{i,j}^t}{b_i^t D}\right|^2 = \frac{\sigma^2}{(b_i^t)^2 D^2} \leq \epsilon\sigma^2, \tag{31}$$

where $\epsilon \in [0,1]$ is reduction rate. According to the observation when $\epsilon \to 0$, (30) is less when $s_i^t = 1$ than when $s_i^t = 0$. Thus, in all rounds, all layers are transmitted and negatively affected by the noise. *2)* In each round t, the communication cost is defined as [43]

$$\text{Communication cost} = \sum_{i=1}^{I} J_i s_i^t, \tag{32}$$

which is the same form as WAE in which some factors are eliminated. This means that the communication cost and WAE increase or decrease simultaneously at roughly the same rate by $\mathbf{s}^t$. Namely, according to the observations, all layers should not be transmitted in all rounds to minimize the negative effect of the noise and the communication cost.

Therefore, to improve the convergence rate and communication efficiency, we defined $r^{1,t}$ as

$$r^{1,t} = (1-\theta)\sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{(b_i^t)^2\lambda^2 D^2} + \theta\sum_{i=1}^{I}\frac{(1-s_i^t\delta)\,G_i^2}{\delta} \times\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right), \tag{33}$$

where $\theta \in [0,1]$ is weight factor.

To determine $\mathbf{b}^t$ and $\mathbf{s}^t$, the PS must consider the transmit power constraint (15) on each device. However, in the middle of (15), the magnitude scaling factor $v_{k,i}^t$ should be estimated because it is unknown to the PS.[3] Thus, under **Assumption 1** and Appendix A, we assumed $\delta = J_i(v_{k,i}^t)^2/\|\mathbf{u}_{k,i}^t\|_2^2$ and then bounded $v_{k,i}^t$ as

$$\begin{aligned}\left(v_{k,i}^t\right)^2 &= \frac{\delta}{J_i}\left\|\mathbf{u}_{k,i}^t\right\|_2^2 \\ &\leq \frac{(2-\delta)\,\lambda^2 G_i^2}{J_i}\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right) \\ &= \left(V_i^{1,t}\right)^2,\end{aligned} \tag{34}$$

where $(V_i^{1,t})^2$ is defined as an achievable upper bound of $(v_{k,i}^t)^2$. Using this, we redefined the transmit power constraint (15) as

$$\left|\frac{b_i^t D_k s_i^t v_{k,i}^t}{h_k^t}\right|^2 \leq \frac{s_i^t D_k^2 (V_i^{1,t})^2 (b_i^t)^2}{(h_k^t)^2} = s_i^t o_{k,i}^{1,t}\left(b_i^t\right)^2, \tag{35}$$

where $o_{k,i}^{1,t}$ is defined by the rest of the factors, except b_i^t and s_i^t. To make (35) satisfy the power constraint,

$$s_i^t o_{k,i}^{1,t}\left(b_i^t\right)^2 \leq P, \tag{36}$$

[3]Among factors, we assume that D_k (i.e., the number of samples on each device) is known to the PS when the FL system initializes the training process and is unchanged until the iteration is over. In addition, we assumed that channel coefficient h_k^t is known to the PS and can be easily estimated through uplink pilot signals, defined in 3GPP standards, transmitted by devices.

using (31) and assuming $t - S_i = M_i$, we made the additional assumption that the maximum non-transmit period M_i is upper bounded layer-wise by a positive constant, that is,

$$\begin{aligned}\mathbb{E}\left[M_i^2\right] &\leq \mathbb{E}\left[\frac{\delta P J_i (h_k^t)^2}{2\,(2-\delta)\,(b_i^t)^2\lambda^2 G_i^2 \mathcal{M}(D_k^2)}\right] \\ &\leq \frac{\epsilon\delta P J_i D^2}{(2-\delta)\,\lambda^2 G_i^2 D_k^2},\end{aligned} \tag{37}$$

where $\mathcal{M}(\cdot)$ is max function. Then, M_i is bounded as

$$M_i \leq \mathcal{M}\left(1, \left\lfloor\sqrt{\frac{\epsilon\delta P J_i D^2}{(2-\delta)\,\lambda^2 G_i^2 \mathcal{M}(D_k^2)}}\right\rfloor\right). \tag{38}$$

Thus, under this assumption, (36) can be satisfied. Based on (33) and (36), the joint optimization problem for SOBAA-EFO is formulated as follows:

$$\begin{aligned}\textbf{P2:}\quad \min_{\mathbf{b}^t,\mathbf{s}^t}\quad & r^{1,t}, && \text{(39a)}\\ \text{s.t.}\quad & s_i^t o_{k,i}^{1,t}\left(b_i^t\right)^2 \leq P, \quad \forall k,i, && \text{(39b)}\\ & t - S_i \leq M_i, \quad \forall i, && \text{(39c)}\\ & \sum_{i=1}^{I} s_i^t \geq 1, && \text{(39d)}\\ & s_i^t \in \{0,1\}, \quad \forall i. && \text{(39e)}\\ & b_i^t > 0, \quad \forall i, && \text{(39f)}\end{aligned}$$

where (39d) is the layer-wise sparsification constraint to ensure that at least one layer is transmitted for convergence performance in each communication round t.

Similarly, we defined $r^{2,t}$ for SOBAA-EFX by dropping the irrelevant terms in (29) and using weight factor θ as

$$r^{2,t} = (1-\theta)\sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{(b_i^t)^2\lambda^2 D^2} + \theta\sum_{i=1}^{I}\left(1-s_i^t\delta\right)G_i^2. \tag{40}$$

Then, based on $\mathbf{u}_k^t = \lambda\mathbf{g}_k^t$ and Cauchy–Schwarz inequality, we bounded $v_{k,i}^t$ as

$$\begin{aligned}\left(v_{k,i}^t\right)^2 &= \frac{1}{J_i^2}\left\|\lambda\mathbf{g}_{k,i}^t\right\|_1^2 \leq \frac{1}{J_i}\left\|\lambda\mathbf{g}_{k,i}^t\right\|_2^2 \\ &\leq \frac{\lambda^2 G_i^2}{J_i} = \left(V_i^{2,t}\right)^2,\end{aligned} \tag{41}$$

where $(V_i^{2,t})^2$ is defined as an achievable upper bound of $(v_{k,i}^t)^2$. Therefore, the joint optimization problem for SOBAA-EFX is formulated as follows:

$$\begin{aligned}\textbf{P3:}\quad \min_{\mathbf{b}^t,\mathbf{s}^t}\quad & r^{2,t}, && \text{(42a)}\\ \text{s.t.}\quad & s_i^t o_{k,i}^{2,t}\left(b_i^t\right)^2 \leq P, \quad \forall k,i, && \text{(42b)}\\ & \sum_{i=1}^{I} s_i^t \geq 1, && \text{(42c)}\\ & s_i^t \in \{0,1\}, \quad \forall i, && \text{(42d)}\\ & b_i^t > 0, \quad \forall i, && \text{(42e)}\end{aligned}$$

where $o_{k,i}^{2,t}$ is defined by $(V_i^{2,t})^2$ instead of $(V_i^{1,t})^2$.

B. Derivation of Computationally Feasible Solution

Because of the combination of the real-valued variable amplitude scaling factor b_i^t and the binary-valued variable sparse masking indicator s_i^t, **P2** is a mixed-integer programming and a non-convex problem. However, b_i^t and s_i^t are not correlated with each other within and between the layers. That outcome means that no matter what given s_i^t, **P2** can become a convex problem that determines the optimal solution of b_i^t. Therefore, we decomposed **P2** into two subproblems for finding b_i^t and s_i^t, respectively.

First, we defined an auxiliary function to determine b_i^t and reformulate the part of **P2** as follows:

$$\textbf{P4:} \quad \min_{\mathbf{b}^t} \quad \mathcal{R}_1(\mathbf{b}^t) = (1-\theta)\sum_{i=1}^{I} \frac{J_i\sigma^2}{(b_i^t)^2\lambda^2 D^2}, \tag{43a}$$

$$\text{s.t.} \quad o_{k,i}^{1,t}\left(b_i^t\right)^2 \leq P, \quad \forall k,i, \tag{43b}$$

$$\text{(39f)}.$$

All b_i^t are not correlated with each other. Thus, we decomposed **P4** into smaller I parallel subproblems as follows:

$$\min_{b_i^t} \quad \mathcal{R}_{1,i}(b_i^t) = (1-\theta)\frac{J_i\sigma^2}{(b_i^t)^2\lambda^2 D^2}. \tag{44}$$

For solving **P4**, we utilized the Lagrange dual method. Therefore, we introduced the multipliers $\mathbf{q}^t = \{q_i^t\}_{i=1}^I$ to constraint (43b) and derived the partial Lagrangian as follows:

$$\mathcal{L}(\mathbf{b}^t, \mathbf{q}^t) = \mathcal{R}_1(\mathbf{b}^t) + \sum_{k=1}^{K}\sum_{i=1}^{I} q_i^t\left(o_{k,i}^{1,t}\left(b_i^t\right)^2 - P\right). \tag{45}$$

Then, the dual problem is derived as follows:

$$\textbf{P5:} \quad \max_{\mathbf{q}^t} \quad \mathcal{D}(\mathbf{q}^t), \tag{46a}$$

$$\text{s.t.} \quad q_i^t \geq 0, \quad \forall i, \tag{46b}$$

where the dual function $\mathcal{D}(\cdot)$ is given by

$$\mathcal{D}(\mathbf{q}^t) = \min_{\mathbf{b}^t} \quad \mathcal{L}(\mathbf{b}^t), \tag{47}$$

$$\text{s.t.} \quad \text{(39f)}.$$

According to the strong duality between problems, the primal problem **P4** is solved by solving its dual problem **P5**. Thus, **P5** was decomposed into smaller I parallel convex subproblems as follows:

$$\min_{b_i^t} \quad \mathcal{L}_i(b_i^t) = \mathcal{R}_{1,i}(b_i^t) + \sum_{k=1}^{K} q_i^t\left(o_{k,i}^{1,t}\left(b_i^t\right)^2 - P\right), \tag{48}$$

which can be solved via efficient convex optimization tools.

Second, we defined another auxiliary function to determine s_i^t and reformulate the part of **P2** as follows:

$$\textbf{P6:} \quad \min_{\mathbf{s}^t} \quad \mathcal{R}_2^1(\mathbf{s}^t) = (1-\theta)\sum_{i=1}^{I}\frac{J_i s_i^t\sigma^2}{(b_i^{t,*})^2\lambda^2 D^2} + \theta\sum_{i=1}^{I}\frac{(1-s_i^t\delta)\,G_i^2}{\delta} \times\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right), \tag{49}$$

$$\text{s.t.} \quad \text{(39c)}, \text{(39d)}, \text{(39e)},$$

where $\mathbf{b}^{t,*} = \{b_i^{t,*}\}_{i=1}^I$ are all determined amplitude scaling factors. Then, **P6** can be solved using the enumeration-based method that enumerates all cases of $\mathbf{s}^t$ and derives the best case among them. However, since this method requires all 2^I cases for the proposed SOBAA-EFO, the result is a computationally infeasible solution. All s_i^t are not correlated with each other; thus, we decomposed **P6** into smaller I parallel subproblems as follows:

$$\min_{s_i^t} \quad \mathcal{R}_{2,i}^1(s_i^t) = (1-\theta)\frac{J_i s_i^t\sigma^2}{(b_i^{t,*})^2\lambda^2 D^2} + \theta\frac{(1-s_i^t\delta)\,G_i^2}{\delta} \times\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right), \tag{50}$$

where if $t - S_i = M_i$ in (39c), the subproblem i is solved to $s_i^t = 1$. Otherwise, this is solved by comparing cases of $s_i^t = 0$ and $s_i^t = 1$, which are expressed as follows:

$$\mathcal{R}_{2,i}^1(0) = \theta\frac{G_i^2}{\delta}\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right) \tag{51}$$

and

$$\mathcal{R}_{2,i}^1(1) = (1-\theta)\frac{J_i\sigma^2}{(b_i^{t,*})^2\lambda^2 D^2} + \theta\frac{(1-\delta)\,G_i^2}{\delta} \times\left((t-S_i)^2 + \frac{(2-\delta)\,M_i^2}{\delta}\right), \tag{52}$$

respectively. Thereafter, the constraint (39d) may not be satisfied. To minimize (49), a layer with smallest (52) should be derived. However, this only derives a specific layer because of differences in the number of parameters and differences in the magnitude of gradients, degrading the convergence rate and learning performance because each layer has a different contribution [44]. Therefore, we derived the layer i^* with the smallest growth rate as follows:

$$i^* = \arg\min_{i} \quad \frac{\mathcal{R}_{2,i}^1(1)}{\mathcal{R}_{2,i}^1(0)}. \tag{53}$$

Then, we obtained all determined sparse masking indicators $\mathbf{s}^{t,*} = \{s_i^{t,*}\}_{i=1}^I$ by setting $s_{i^*}^t = 1$.

The computational complexity of solving **P5** and **P6** is both $\mathcal{O}(I)$ because the computational complexity of solving smaller parallel subproblems is scaled with I. Thus, the computational complexity of the proposed method to solve **P2**, which solves **P5** and **P6** sequentially, is $\mathcal{O}(I)$, which is scaled linearly with I. Therefore, the proposed method determines a solution with the computationally feasible complexity. Moreover, the proposed method is expressed as the pseudo-code in **Algorithm 1**, where $\mathbf{o}^{1,t}$ denotes all $o_{k,i}^{1,t}$ for all k and i.

In each communication round t, the PS determines $\mathbf{b}^{t,*}$ and $\mathbf{s}^{t,*}$ using **Algorithm 1** and broadcasts them to all devices. Thus, given $\mathbf{b}^{t,*}$ and $\mathbf{s}^{t,*}$, each device derives $\ddot{\mathbf{u}}_k^t$ based on (10) and then transmits it based on (16).

For solving **P3**, we also used (43a) as the auxiliary function and defined another auxiliary function as follows:

$$\mathcal{R}_2^2(\mathbf{s}^t) = (1-\theta)\sum_{i=1}^{I}\frac{J_i s_i^t\sigma^2}{(b_i^{t,*})^2\lambda^2 D^2} + \theta\sum_{i=1}^{I}\left(1-s_i^t\delta\right)G_i^2, \tag{54}$$

TABLE II: Experiment parameter setup.

Parameter	Value		Parameter		Value					
	MNIST dataset	CIFAR-10 dataset			MLP classifier		4NN classifier		5NN classifier	
					EFO	EFX	EFO	EFX	EFO	EFX
Number of devices K	$[5, 25]$		Approximation factor δ		0.02	0.30	0.01	0.10	0.01	0.07
Learning rate λ	0.01		Weight factor θ for (K, σ^2)	$(5, 10^{-4})$	5.e-8	5.e-5	5.e-8	5.e-5	5.e-8	5.e-5
Maximum communication round T	200	500		$(10, 10^{-4})$	5.e-8	5.e-6	5.e-8	5.e-5	5.e-8	5.e-5
				$(15, 10^{-4})$	5.e-8	5.e-6	5.e-8	5.e-6	5.e-8	5.e-6
Maximum transmit power P	10			$(20, 10^{-4})$	5.e-8	5.e-6	5.e-8	5.e-8	5.e-8	5.e-6
				$(25, 10^{-4})$	5.e-8	5.e-6	5.e-8	5.e-8	5.e-8	5.e-6
Noise power σ^2	$[1, 10^{-4}]$			$(25, 10^{-3})$	5.e-8	5.e-5	5.e-8	5.e-8	5.e-8	5.e-5
				$(25, 10^{-2})$	5.e-8	5.e-4	5.e-8	5.e-8	5.e-8	5.e-4
Reduction rate ϵ	5.e-7			$(25, 10^{-1})$	5.e-8	5.e-3	5.e-8	5.e-8	5.e-8	5.e-3
				$(25, 1)$	5.e-8	5.e-2	5.e-8	5.e-8	5.e-7	5.e-2

Algorithm 1 Decomposition-Based Feasible Solution

Input: $\mathbf{o}^{1,t}$.
Output: $\mathbf{b}^{t,*}$, $\mathbf{s}^{t,*}$.
1: **Initialize** $\mathbf{b}^t$, $\mathbf{s}^t$, $\mathbf{q}^t$.
2: **for** each layer $i \in \{1, \ldots, I\}$ **in parallel do**
3: Obtain b_i^t by solving (48).
4: Obtain s_i^t by solving (50).
5: **end for**
6: **if** (39d) is not satisfied **then**
7: Obtain i^* by solving (53).
8: Set $s_{i^*}^t = 1$.
9: **end if**
10: **return** $\mathbf{b}^{t,*}$, $\mathbf{s}^{t,*}$.

where the remaining process is the same as the solving process for **P2**, thus it is omitted.

V. Experimental Results

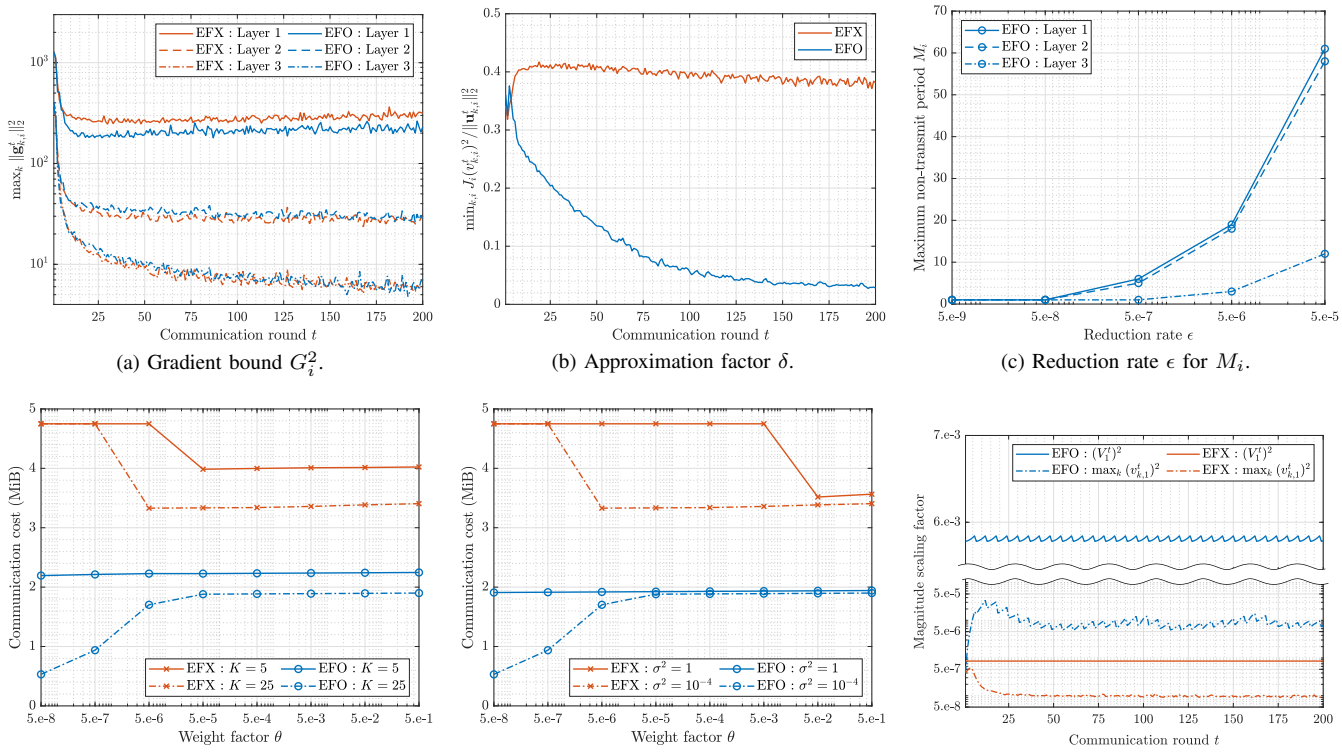
This section evaluates the learning performance and communication efficiency of the proposed SOBAA-based FLOA system. We considered the FLOA system with one server and $K = 25$ devices, which were referenced from [25], [26], and considered maximum transmission power $P = 10$ mW and noise power $\sigma^2 = 10^{-4}$ mW, which were referenced from [29]. We compared the proposed SOBAA with OBDA [27] and EFOBDA [28], In terms of the gradient compression scheme, the previous work OBDA [27] is a method that adopts unscaled one-bit quantization, but it does not adopt sparsification and error-feedback mechanism. The latter work EFOBDA [28] is a more advanced method that adopts unscaled one-bit quantization and error-feedback mechanism, but it does not adopt sparsification. The proposed SOBAA-EFO/EFX is the method that adopts layer-wise scaled one-bit quantization and layer-wise sparsification with/without the error-feedback mechanism.

As an application of this experimental evaluation, we considered image classification which is a fundamental task in vision recognition, using well-known MNIST and CIFAR-10 datasets. The MNIST dataset consists of 10 classes with black-and-white handwritten digits ranging from 0 to 9 that are one color with 28×28 pixels, where $D = 60000$ labeled training samples and 10000 test samples are available. The CIFAR-10 dataset consists of 10 mutually exclusive classes with colorful objects or animals that are RGB color with 32×32 pixels, where $D = 50000$ labeled training samples and 10000 test samples are available.

In the numerical experiments, we considered two cases of partitioning the dataset over 25 devices: *1) an i.i.d. case for both MNIST and CIFAR-10 datasets*, where the samples are shuffled and partitioned into 25 devices, each receiving $D_k = D/25$ samples, and *2) a non-i.i.d. case for only MNIST dataset*, where the samples are sorted by digit label divided into 200 shards with 300 samples and then assigned to the 25 devices, each receiving 8 shards.

In particular, for the i.i.d. MNIST dataset, the classifier is implemented using a simple MLP that consists of two hidden layers with 200 units each using the rectified linear units (ReLU) activation function (i.e., $I = 3$ and $J = 199210$). For the non-i.i.d. MNIST dataset, the classifier is implemented using a CNN consisting of two 5×5 convolutional layers with ReLU activation (the first with 32 channels, and the second with 64 channels, each followed by 2×2 max-pooling), a fully connected layer with 512 units and ReLU activation, and a final softmax output layer (i.e., $I = 4$ and $J = 582026$), which is denoted as 4NN. For the i.i.d. CIFAR-10 dataset, the classifier is implemented using a CNN consisting of two convolutional layers, two fully connected layers, and a linear transformation layer to produce logits (i.e., $I = 5$ and $J = 940362$), which is denoted as 5NN. For the sake of reliability, we used the same classifiers implemented for each dataset in prior FL-related works. That is, the MLP and 5NN are referred from [2], and the 4NN is referred from [27].

All environments were implemented using the Matlab optimization toolbox and Python programming language including Pytorch library on an Intel(R) Core i9-13900K machine with NVIDIA Geforce RTX 4080 and 32 GB memory, and all methods were evaluated in the same environment. By observing the parameters where all combinations of classifier and method converge in test accuracy, the learning rate λ is set to 0.1, and the maximum communication round T of the MNIST and CIFAR-10 datasets are set to 200 and 500, respectively.



(a) Gradient bound G_i^2. (b) Approximation factor δ. (c) Reduction rate ϵ for M_i.

(d) Weight factor θ for $(K, \sigma^2 = 10^{-4})$. (e) Weight factor θ for $(K = 25, \sigma^2)$. (f) Magnitude scaling factor.

Fig. 2: Preliminary experiments for parameter setup on MLP classifier with i.i.d. MNIST dataset.

Also, we performed preliminary experiments for parameter setup, for example, on MLP classifier with i.i.d. MNIST dataset. Fig. 2 shows experiment results. Assuming an error-free channel (i.e., $h_k^t = 1$ and $\mathbf{z}^t = \mathbf{0}$), both EFO and EFX do not perform layer-wise sparsification at all communication rounds. Therefore, we set G_i^2 by $\max_k \|\mathbf{g}_{k,i}^t\|_2^2$ in (25) using Fig. 2a. In addition, we set δ by $\min_{k,i} J_i(v_{k,i}^t)^2/\|\mathbf{u}_{k,i}^t\|_2^2$ in (20) using Fig. 2b. Moreover, we set ϵ based on (38) so that all M_i becomes smaller than or equal to 10 (i.e., each layer being transmitted at least once within 10 rounds) as shown in Fig. 2c.

On the other hand, under a fading channel, we set the weight factor θ for each device population and each network condition using Figs. 2d and 2e, respectively. This plots the required communication costs based on (32) to achieve a 0.95 target test accuracy. When the target test accuracy is not achieved, the communication cost is calculated as $J \times T$. Regardless of θ, EFO is always lower than EFX due to the error-feedback mechanism. For both EFO and EFX, as θ decreases, the first term on the RHS of (33) and (40) increases, resulting in infrequent transmission of layers and lower communication costs. However, EFX fails to achieve the 0.95 target test accuracy when θ is significantly lower than a certain threshold, resulting in dramatically increased communication costs. In FLOA, because all devices transmit simultaneously over the same time-frequency resources, the communication cost does not increase linearly with the number of devices, and it is the same as the communication cost of a single device [43]. On the other hand, as the number of devices increases, the learning performance converges faster (because of increased training samples, $D = D_k \times K$). That is, the required number of communication rounds to the convergence decreases, and it brings reduced communication costs as shown in Fig. 2d. Moreover, it also reduces the thresholds of θ since the positive effect from the increased number of training samples mitigates the negative effect from the infrequent transmission of layers. Similarly, Fig. 2e shows that as a noise power decreases, the learning performance converges faster (because of a less distorted signal) and the required communication cost is decreased. Thus, for each combination (K, σ^2), we selected an appropriate θ that requires the least communication cost to achieve the 0.95 target test accuracy. To set parameters on the other classifiers and datasets, we performed similar preliminary experiments and then determined the parameters. Table II lists the experiment parameter setup.

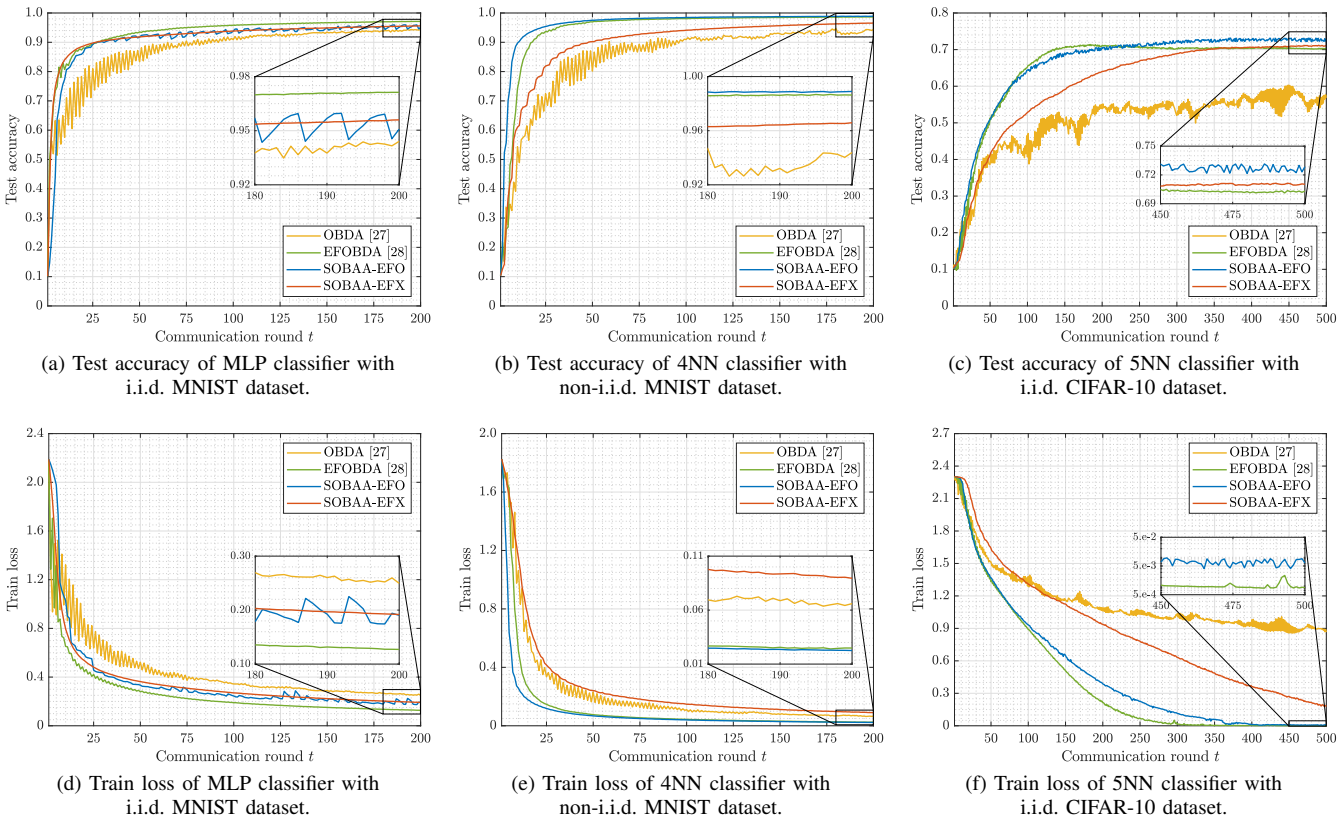
Based on (8), (34), and (41), we measured $\max_k (v_{k,1}^t)^2$, $(V_1^{1,t})^2$, and $(V_1^{2,t})^2$. As shown in Fig. 2f, the transmitted magnitude scaling factor is always lower than the estimated for both EFO and EFX in all communication rounds, thus we showed that the estimation operation is well performed.

A. *Evaluation of Learning Performance*

The effectiveness of the SOBAA-EFO and SOBAA-EFX are evaluated for the MLP with the i.i.d. MNIST, the 4NN with the non-i.i.d. MNIST, and the 5NN with the i.i.d. CIFAR-10. Table III listed the maximum test accuracy, minimum train loss, and ranking in the error-free channels. Fig. 3 plotted the

TABLE III: Evaluation of learning performance on error-free channels.

Classifier and dataset	Maximum test accuracy (Ranking)				Minimum train loss (Ranking)			
	OBDA [27]	EFOBDA [28]	SOBAA-EFO	SOBAA-EFX	OBDA [27]	EFOBDA [28]	SOBAA-EFO	SOBAA-EFX
MLP classifier with i.i.d. MNIST dataset	0.94613 (4)	0.97213 (1)	0.96893 (2)	0.95767 (3)	0.24557 (4)	0.12638 (1)	0.13544 (2)	0.18245 (3)
4NN classifier with non-i.i.d. MNIST dataset	0.97660 (4)	0.98843 (2)	0.98950 (1)	0.98680 (3)	0.03967 (4)	0.02307 (2)	0.02121 (1)	0.03374 (3)
5NN classifier with i.i.d. CIFAR-10 dataset	0.62230 (4)	0.72470 (3)	0.73187 (1)	0.72660 (2)	0.81729 (4)	0.00082 (2)	0.00081 (1)	0.00748 (3)



(a) Test accuracy of MLP classifier with i.i.d. MNIST dataset.

(b) Test accuracy of 4NN classifier with non-i.i.d. MNIST dataset.

(c) Test accuracy of 5NN classifier with i.i.d. CIFAR-10 dataset.

(d) Train loss of MLP classifier with i.i.d. MNIST dataset.

(e) Train loss of 4NN classifier with non-i.i.d. MNIST dataset.

(f) Train loss of 5NN classifier with i.i.d. CIFAR-10 dataset.

Fig. 3: Evaluation of learning performance on fading channels.

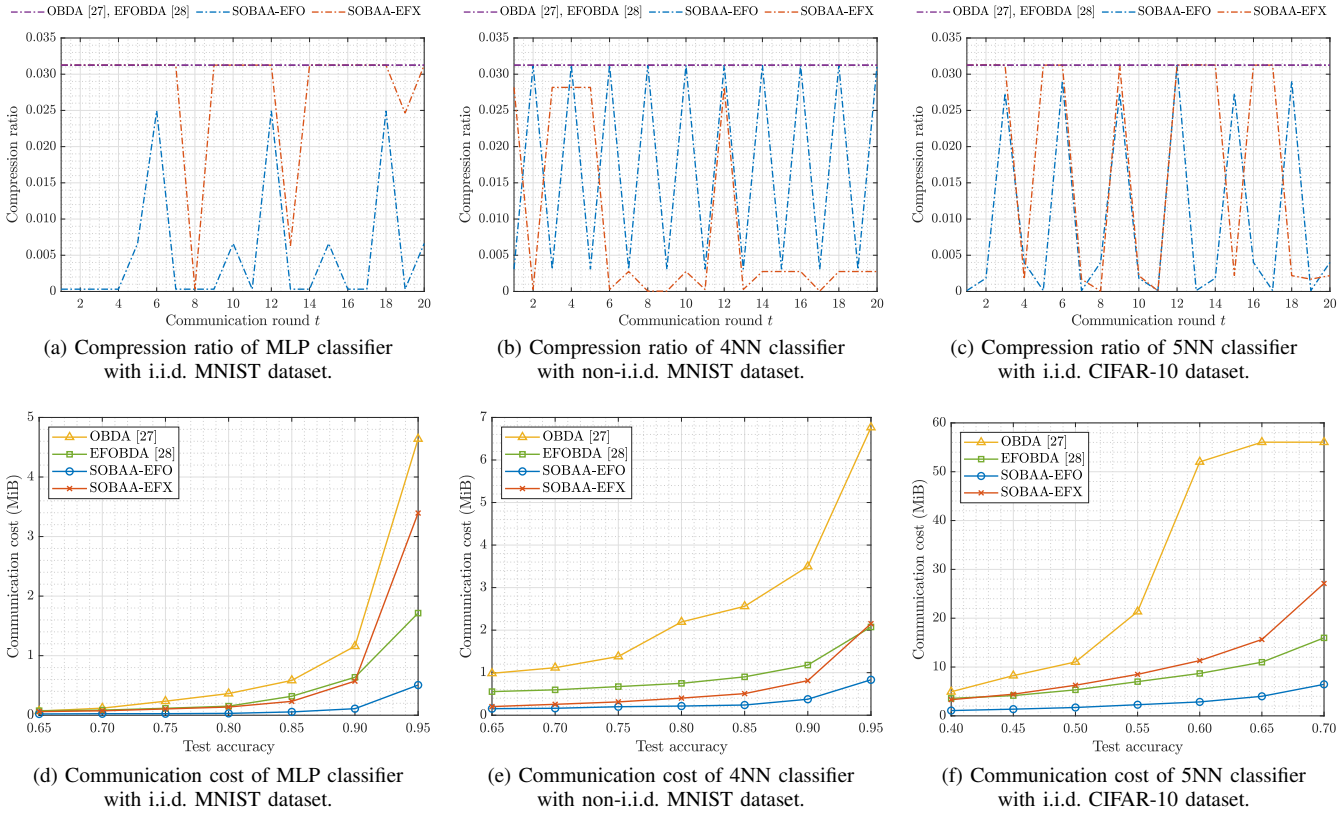
test accuracy and train loss according to the communication rounds on the fading channels.

For the MLP with the i.i.d. MNIST, in the error-free channels, SOBAA-EFO achieves better performance than SOBAA-EFX and OBDA. However, it achieves slightly lower performance than EFOBDA, which is because aggregating only the sign (EFOBDA) may achieve better learning performance than aggregating both the sign and the magnitude (SOBAA-EFO) depending on the dataset and model [8]. However, the test accuracy and train loss gap are significantly small (i.e., 0.003 and 0.009, respectively). In fading channels, SOBAA-EFO achieves the same ranking as shown in Figs. 3a and 3d.

For the 4NN with the non-i.i.d. MNIST, in error-free channels, SOBAA-EFO achieves the highest learning performance, followed in order by EFOBDA, SOBAA-EFX, and OBDA. In fading channels, SOBAA-EFO achieves the same ranking as shown in Figs. 3b and 3e, but SOBAA-EFX has a little bit higher train loss than OBDA due to layer-wise sparsification.

For the 5NN with the i.i.d. CIFAR-10, in error-free channels, SOBAA-EFO achieves the highest learning performance, followed in order by SOBAA-EFX, EFOBDA, and OBDA with test accuracy, but followed in order by EFOBDA, SOBAA-EFX, and OBDA with train loss. In fading channels, SOBAA-EFO achieves the same ranking with test accuracy as shown in Fig. 3c, but it has a little bit higher train loss than EFOBDA due to layer-wise sparsification as shown in Fig. 3f.

Therefore, the proposed SOBAA showed a better learning performance as the classifier and dataset became more complex. Notably, for the 5NN with the i.i.d. CIFAR-10, SOBAA-EFX achieves better test accuracy than EFOBDA.



(a) Compression ratio of MLP classifier with i.i.d. MNIST dataset.

(b) Compression ratio of 4NN classifier with non-i.i.d. MNIST dataset.

(c) Compression ratio of 5NN classifier with i.i.d. CIFAR-10 dataset.

(d) Communication cost of MLP classifier with i.i.d. MNIST dataset.

(e) Communication cost of 4NN classifier with non-i.i.d. MNIST dataset.

(f) Communication cost of 5NN classifier with i.i.d. CIFAR-10 dataset.

Fig. 4: Evaluation of communication efficiency.

B. Evaluation of Communication Efficiency

In each communication round t, the compression ratio is defined as [45]

$$\text{Compression ratio} = \frac{1}{32J}\sum_{i=1}^{I} J_i s_i^t, \tag{55}$$

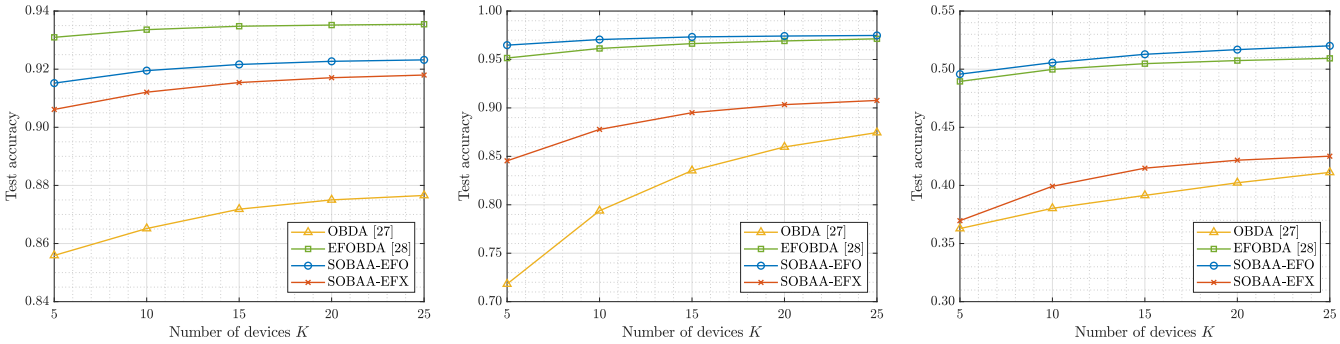
where 32 denotes the number of bits by FloatTensor of the Pytorch library. Based on (55), Figs. 4a, 4b, and 4c showed that the proposed SOBAA optimized the gradient compression ratio along with the communication round compared to conventional methods. Unlike OBDA and EFOBDA, which show a fixed compression ratio of 1/32 across all classifiers and all datasets, SOBAA-EFX shows a different compression ratio according to the communication round.

Then, we compared the proposed SOBAA with the conventional schemes in terms of the trade-off between the communication cost and test accuracy (i.e., the required communication cost to obtain a target test accuracy). When the target accuracy is not achieved, the communication cost is calculated as $J \times T$. As presented in Figs. 4d, 4e and 4f, the communication cost of OBDA increases dramatically as the test accuracy increases, exhibiting the worst trade-off. Moreover, OBDA may fail to achieve a 0.95, 0.95, and 0.6 test accuracy on the MLP, 4NN, and 5NN, respectively. In addition, the trade-off for EFOBDA is improved much more than that for OBDA because EFOBDA achieved a faster convergence rate than OBDA due to the error-feedback mechanism and power control optimization, resulting in lower required communication costs than OBDA. Notably, SOBAA-EFX achieved a better trade-off than EFOBDA except for 0.95 target test accuracy on the MLP and 4NN. However, SOBAA-EFO provides the best trade-off. For a 0.95 target test accuracy, compared to EFOBDA, SOBAA-EFO has 70.5% and 59.8% reduced communication costs on the MLP and 4NN, respectively. In addition, for a 0.7 target test accuracy, SOBAA-EFO has 59.6% reduced communication cost on the 5NN, compared to EFOBDA.

C. Effect of Device Population

In Fig. 5, we set the maximum communication round T to 50 and compared the proposed SOBAA with other methods in terms of the test accuracy according to the number of devices K. For all methods, as K decreases, the test accuracy decreases because the small K results in the small training samples that can be utilized. In particular, it degrades exponentially fast when K goes below a certain threshold.

Furthermore, we showed the robustness of the proposed methods by comparing the test accuracy with $K = 25$ and 5. When the number of devices k decreases from 25 to 5, for the MLP with i.i.d. MNIST, the test accuracy of the proposed SOBAA-EFO decreases only by 0.006, and it can be seen that SOBAA-EFO is more robust than OBDA and SOBAA-EFX, and a little bit less robust than EFOBDA. For the 4NN with non-i.i.d. MNIST and the 5NN with CIFAR-10, the test accuracy of the proposed SOBAA-EFO decreases only by 0.01, and it can be seen that SOBAA-EFO is more robust than OBDA, EFOBDA, and SOBAA-EFX.

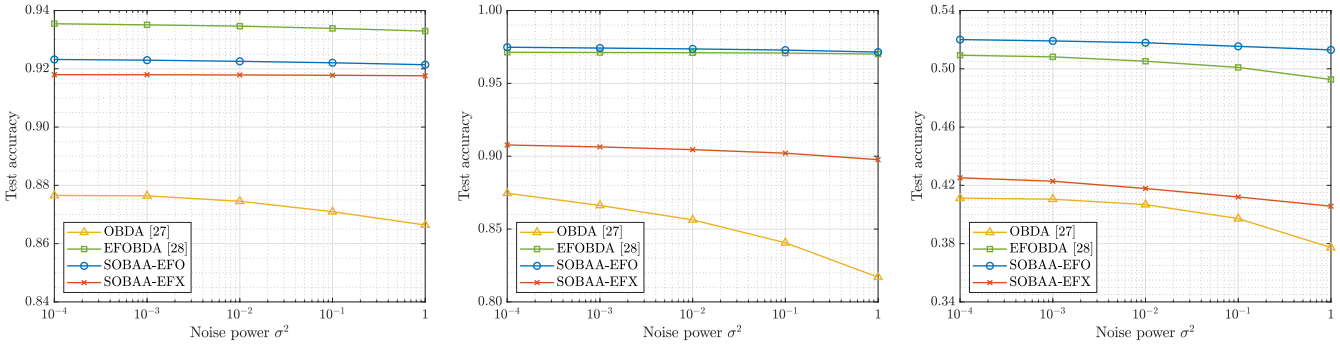


(a) Test accuracy of MLP classifier with i.i.d. MNIST dataset.

(b) Test accuracy of 4NN classifier with non-i.i.d. MNIST dataset.

(c) Test accuracy of 5NN classifier with i.i.d. CIFAR-10 dataset.

Fig. 5: Effect of device population.



(a) Test accuracy of MLP classifier with i.i.d. MNIST dataset.

(b) Test accuracy of 4NN classifier with non-i.i.d. MNIST dataset.

(c) Test accuracy of 5NN classifier with i.i.d. CIFAR-10 dataset.

Fig. 6: Effect of network condition.

D. Effect of Network Condition

In Fig. 6, we set the maximum communication round T to 50 and compared the proposed methods with other methods in terms of the test accuracy according to the noise power σ^2. For all methods, as σ^2 increases, the test accuracy decreases because the signal distortion by noise (i.e., the negative effect of noise) is increased, and its degradation goes exponentially fast when σ^2 goes beyond a certain threshold. Moreover, for both SOBAA-EFO and SOBAA-EFX, the WCE almost does not change while the WAE increases proportionally degrading the convergence rate and learning performance. However, the proposed SOBAA can mitigate this degeneration by adopting an appropriate weight factor θ.

Thus, we showed the robustness of the proposed methods by comparing the test accuracy with $\sigma^2 = 10^{-4}$ and 1. When the noise power σ^2 increases from 10^{-4} to 1, for the MLP with i.i.d. MNIST, the test accuracy of the proposed SOBAA-EFX decreases only by 0.0004, and it can be seen that SOBAA-EFX is more robust than OBDA, EFOBDA, and SOBAA-EFO. For the 4NN with non-i.i.d. MNIST, the test accuracy of the proposed SOBAA-EFO decreases only by 0.003, and it can be seen that SOBAA-EFO is more robust than OBDA and SOBAA-EFX, but a little bit less robust than EFOBDA. For the 5NN with CIFAR-10, the test accuracy of the proposed SOBAA-EFO decreases only by 0.007, and it can be seen that SOBAA-EFO is more robust than OBDA, EFOBDA, and SOBAA-EFX.

VI. Conclusion

For communication-efficient FL, a framework for realizing distributed machine learning, many over-the-air transmission schemes have been proposed. However, in FLOA, some problems remain, such as a slow convergence rate caused by compression and aggregation errors. This work addressed this problem by proposing the SOBAA scheme, incorporating layer-wise scaled one-bit quantization, layer-wise sparsification, error-feedback mechanism, and power control into FLOA. We derived the upper bound of the convergence rate of the proposed method as a closed-form expression. From this, we identified the relationship between the convergence rate and aggregation and compression errors. Based on this relationship, we determined a compression ratio and transmit power that minimizes compression and aggregation errors, leading to the fastest convergence. The optimization problem was decomposed and solved with polynomial complexity. From this solution, we characterized the trade-off between learning performance and the communication cost. Through extensive simulations with the i.i.d. and non-i.i.d. datasets, we confirmed that the proposed SOBAA approach provides an enhanced trade-off between test accuracy and communication

cost and an enhanced convergence rate compared to state-of-the-art approaches. The proposed approach is more effective with complex datasets and learning models.

APPENDIX

A. Proof of *Lemma 1*

According to (7), (11), and **Assumption 1**, the compression error $\mathbf{c}_k^{t+1}$ is bounded layer-wise as

$$\left\|\mathbf{c}_{k,i}^{t+1}\right\|_2^2 = \left\|\mathcal{C}(\mathbf{u}_{k,i}^t) - \mathbf{u}_{k,i}^t\right\|_2^2 \leq \left(1 - s_i^t\delta\right)\left\|\mathbf{u}_{k,i}^t\right\|_2^2 = \left(1 - s_i^t\delta\right)\left\|\sum_{\tau=S_i+1}^{t}\lambda\mathbf{g}_k^\tau + \mathbf{c}_k^{S_i+1}\right\|_2^2. \tag{56}$$

Thus, we have the recurrence relation on the bound. Using the Peter-Paul inequality with any $\eta > 0$, Jensen's inequality, and **Assumption 4**, we obtain

$$\begin{aligned}\left\|\mathbf{c}_{k,i}^{t+1}\right\|_2^2 &\leq \left(1 - s_i^t\delta\right)\left\|\sum_{\tau=S_i+1}^{t}\lambda\mathbf{g}_k^\tau + \mathbf{c}_k^{S_i+1}\right\|_2^2 \\ &\leq \left(1 - s_i^t\delta\right)(1 + 1/\eta)\left\|\sum_{\tau=S_i+1}^{t}\lambda\mathbf{g}_k^\tau\right\|_2^2 + \left(1 - s_i^t\delta\right)(1+\eta)\left\|\mathbf{c}_k^{S_i+1}\right\|_2^2 \\ &\leq \left(1 - s_i^t\delta\right)(1 + 1/\eta)(t - S_i)^2\lambda^2 G_i^2 + \left(1 - s_i^t\delta\right)(1+\eta)\left\|\mathbf{c}_k^{S_i+1}\right\|_2^2. \end{aligned} \tag{57}$$

To solve the recurrence relation, we defined the memory $m_i \in \{0, \ldots, S_i\}$ that stores all S_i from the past to the present. By applying (56) and (57) to all communication rounds stored in m_i, we obtain

$$\begin{aligned}\mathbb{E}\left\|\mathbf{c}_{k,i}^{t+1}\right\|_2^2 &\leq \left(1 - s_i^t\delta\right)(1 + 1/\eta)(t - S_i)^2\lambda^2 G_i^2 + \left(1 - s_i^t\delta\right)(1+\eta)\,\mathbb{E}\left\|\mathbf{c}_k^{S_i+1}\right\|_2^2 \\ &\leq \left(1 - s_i^t\delta\right)(1 + 1/\eta)(t - S_i)^2\lambda^2 G_i^2 + \left(1 - s_i^t\delta\right)(1+\eta)(1-\delta)(1+1/\eta)\lambda^2 G_i^2 \\ &\quad\times\sum_{n=2}^{N_i}\left((1-\delta)(1+\eta)\right)^{N_i-n}\left(m_{i,n} - m_{i,n-1}\right)^2 \\ &\leq \left(1 - s_i^t\delta\right)(1 + 1/\eta)(t - S_i)^2\lambda^2 G_i^2 + \left(1 - s_i^t\delta\right)(1+\eta)(1-\delta)(1+1/\eta)\lambda^2 M_i^2 G_i^2 \\ &\quad\times\sum_{n=2}^{\infty}\left((1-\delta)(1+\eta)\right)^{n-2} \\ &= \left(1 - s_i^t\delta\right)(1 + 1/\eta)(t - S_i)^2\lambda^2 G_i^2 + \frac{\left(1 - s_i^t\delta\right)(1+\eta)(1-\delta)(1+1/\eta)\lambda^2 M_i^2 G_i^2}{1-(1-\delta)(1+\eta)} \\ &\leq \frac{\left(1 - s_i^t\delta\right)(2-\delta)\lambda^2 G_i^2}{\delta} \times\left((t - S_i)^2 + \frac{(2-\delta)M_i^2}{\delta}\right), \end{aligned} \tag{58}$$

where $n \in \{1, \ldots, N_i\}$ is the address n. In addition, we set $\eta = \frac{\delta}{2(1-\delta)}$. Then, according to Jensen's inequality,

$$\begin{aligned}\mathbb{E}\left\|\frac{1}{D}\sum_{k=1}^{K} D_k\mathbf{c}_k^{t+1}\right\|_2^2 &\leq \frac{1}{D}\sum_{k=1}^{K} D_k\mathbb{E}\left\|\mathbf{c}_k^{t+1}\right\|_2^2 \\ &= \frac{1}{D}\sum_{k=1}^{K} D_k\sum_{i=1}^{I}\mathbb{E}\left\|\mathbf{c}_{k,i}^{t+1}\right\|_2^2 \\ &\leq \sum_{i=1}^{I}\frac{\left(1 - s_i^t\delta\right)(2-\delta)\lambda^2 G_i^2}{\delta} \times\left((t - S_i)^2 + \frac{(2-\delta)M_i^2}{\delta}\right), \end{aligned} \tag{59}$$

which completes the proof. ■

B. Proof of *Theorem 1*

Following the widely adopted strategy in this proof under **Assumption 3**, we considered the non-compression scenario that sets the actual parameter vector $\mathbf{w}^t$ to be an approximation to the ideal parameter vector $\hat{\mathbf{w}}^t$, defined as

$$\hat{\mathbf{w}}^t = \mathbf{w}^t - \frac{1}{D}\sum_{k=1}^{K} D_k\mathbf{c}_k^t. \tag{60}$$

Then, we have **Lemma 3**.

Lemma 3. The ideal parameter vector $\hat{\mathbf{w}}^t$ is updated as

$$\hat{\mathbf{w}}^{t+1} = \hat{\mathbf{w}}^t - \lambda\mathbf{g}^t - \mathbf{a}^t. \tag{61}$$

Proof: See Appendix C. ■

According to **Assumption 3**, in each communication round, the global loss function $\mathcal{F}(\cdot)$ with $\hat{\mathbf{w}}^{t+1}$ is bounded by

$$\mathcal{F}(\hat{\mathbf{w}}^{t+1}) \leq \mathcal{F}(\hat{\mathbf{w}}^t) + \left(\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right)^\top\nabla\mathcal{F}(\hat{\mathbf{w}}^t) + \frac{L}{2}\left\|\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right\|_2^2. \tag{62}$$

By **Assumption 4**,

$$\begin{aligned}&\mathbb{E}\left[\mathcal{F}(\hat{\mathbf{w}}^{t+1})\right] \\ &\leq \mathcal{F}(\hat{\mathbf{w}}^t) + \mathbb{E}\left[\left(\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right)\right]^\top\nabla\mathcal{F}(\hat{\mathbf{w}}^t) + \frac{L}{2}\mathbb{E}\left\|\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right\|_2^2 \\ &= \mathcal{F}(\hat{\mathbf{w}}^t) - \left(\lambda\mathbb{E}\left[\mathbf{g}^t\right] + \mathbb{E}\left[\mathbf{a}^t\right]\right)^\top\nabla\mathcal{F}(\hat{\mathbf{w}}^t) + \frac{L}{2}\mathbb{E}\left\|\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right\|_2^2 \\ &= \mathcal{F}(\hat{\mathbf{w}}^t) - \lambda\nabla\mathcal{F}(\mathbf{w}^t)^\top\nabla\mathcal{F}(\hat{\mathbf{w}}^t) + \frac{L}{2}\mathbb{E}\left\|\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right\|_2^2 \\ &= \mathcal{F}(\hat{\mathbf{w}}^t) + \lambda\nabla\mathcal{F}(\mathbf{w}^t)^\top\left(\nabla\mathcal{F}(\mathbf{w}^t) - \nabla\mathcal{F}(\hat{\mathbf{w}}^t)\right) - \lambda\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \frac{L}{2}\mathbb{E}\left\|\hat{\mathbf{w}}^{t+1} - \hat{\mathbf{w}}^t\right\|_2^2. \end{aligned} \tag{63}$$

On the RHS of (63), the second term is bounded by the Peter-Paul inequality with any $\rho > 0$ as

$$\lambda\nabla\mathcal{F}(\mathbf{w}^t)^\top\left(\nabla\mathcal{F}(\mathbf{w}^t) - \nabla\mathcal{F}(\hat{\mathbf{w}}^t)\right) \leq \frac{\lambda\rho}{2}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \frac{\lambda}{2\rho}\left\|\nabla\mathcal{F}(\mathbf{w}^t) - \nabla\mathcal{F}(\hat{\mathbf{w}}^t)\right\|_2^2, \tag{64}$$

where the second term on the RHS is bounded by **Assumption 3** and **Lemma 1** as

$$\begin{aligned}
&\frac{\lambda}{2\rho}\left\|\nabla\mathcal{F}(\mathbf{w}^t)-\nabla\mathcal{F}(\hat{\mathbf{w}}^t)\right\|_2^2\\
&\leq\frac{\lambda L^2}{2\rho}\left\|\mathbf{w}^t-\hat{\mathbf{w}}^t\right\|_2^2\\
&=\frac{\lambda L^2}{2\rho}\left\|\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^t\right\|_2^2\\
&\leq\sum_{i=1}^{I}\frac{\left(1-s_i^{t-1}\delta\right)(2-\delta)\lambda^3L^2G_i^2}{2\rho\delta}\\
&\quad\times\left((t-1-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right).
\end{aligned}\tag{65}$$

Therefore, the second term on the RHS of (63) is bounded by

$$\begin{aligned}
&\lambda\nabla\mathcal{F}(\mathbf{w}^t)^\top\left(\nabla\mathcal{F}(\mathbf{w}^t)-\nabla\mathcal{F}(\hat{\mathbf{w}}^t)\right)\\
&\leq\frac{\lambda\rho}{2}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2+\sum_{i=1}^{I}\frac{\left(1-s_i^{t-1}\delta\right)(2-\delta)\lambda^3L^2G_i^2}{2\rho\delta}\\
&\quad\times\left((t-1-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right).
\end{aligned}\tag{66}$$

Then, the last term on the RHS of (63) is bounded by Jensen's inequality, **Assumption 4**, and **Lemma 2** as

$$\begin{aligned}
&\frac{L}{2}\mathbb{E}\left\|\hat{\mathbf{w}}^{t+1}-\hat{\mathbf{w}}^t\right\|_2^2\\
&=\frac{\lambda^2L}{2}\mathbb{E}\left\|\mathbf{g}^t\right\|_2^2+\frac{L}{2}\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2\\
&=\frac{\lambda^2L}{2}\mathbb{E}\left\|\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{g}_k^t\right\|_2^2+\frac{L}{2}\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2\\
&\leq\frac{\lambda^2L}{2D}\sum_{k=1}^{K}D_k\mathbb{E}\left\|\mathbf{g}_k^t\right\|_2^2+\frac{L}{2}\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2\\
&\leq\frac{\lambda^2LG^2}{2}+\frac{L}{2}\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2\\
&=\frac{\lambda^2LG^2}{2}+\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2L}{2(b_i^t)^2D^2}.
\end{aligned}\tag{67}$$

By inserting (66) and (67) into (63), we obtain

$$\begin{aligned}
\mathbb{E}\left[\mathcal{F}(\hat{\mathbf{w}}^{t+1})\right]&\leq\mathcal{F}(\hat{\mathbf{w}}^t)-\lambda\left(1-\frac{\rho}{2}\right)\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2\\
&\quad+\sum_{i=1}^{I}\frac{\left(1-s_i^{t-1}\delta\right)(2-\delta)\lambda^3L^2G_i^2}{2\rho\delta}\\
&\quad\times\left((t-1-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right)\\
&\quad+\frac{\lambda^2LG^2}{2}+\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2L}{2(b_i^t)^2D^2}.
\end{aligned}\tag{68}$$

According to **Assumption 2**, we rearrange the terms and take the average over t on both sides as

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2&\leq\frac{2\left(\mathcal{F}(\mathbf{w}^1)-\mathcal{F}(\mathbf{w}^*)\right)}{(2-\rho)\lambda T}+\frac{\lambda LG^2}{2-\rho}\\
&\quad+\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2L}{(2-\rho)(b_i^t)^2\lambda D^2T}\\
&\quad+\sum_{t=1}^{T-1}\sum_{i=1}^{I}\frac{(1-s_i^t\delta)(2-\delta)\lambda^2L^2G_i^2}{(2-\rho)\rho\delta T}\\
&\quad\times\left((t-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right),
\end{aligned}\tag{69}$$

where L and ρ are set to $1/\lambda$ and $2-\delta$, respectively, for a simpler expression without compromising generality, then

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left\|\nabla\mathcal{F}(\mathbf{w}^t)\right\|_2^2&\leq\frac{2\left(\mathcal{F}(\mathbf{w}^1)-\mathcal{F}(\mathbf{w}^*)\right)}{\delta\lambda T}+\frac{G^2}{\delta}\\
&\quad+\frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{J_is_i^t\sigma^2}{\delta(b_i^t)^2\lambda^2D^2}\\
&\quad+\frac{1}{T}\sum_{t=1}^{T-1}\sum_{i=1}^{I}\frac{(1-s_i^t\delta)G_i^2}{\delta^2}\\
&\quad\times\left((t-S_i)^2+\frac{(2-\delta)M_i^2}{\delta}\right),
\end{aligned}\tag{70}$$

which completes the proof. ■

C. Proof of *Lemma 3*

By sequentially inserting (18), (17), (11), and (7) into (60), we obtain

$$\begin{aligned}
\hat{\mathbf{w}}^{t+1}&=\mathbf{w}^{t+1}-\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^{t+1}\\
&=\mathbf{w}^t-\tilde{\mathbf{u}}^t-\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^{t+1}\\
&=\mathbf{w}^t-\ddot{\mathbf{u}}^t-\mathbf{a}^t-\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^{t+1}\\
&=\mathbf{w}^t-\mathbf{u}^t-\mathbf{a}^t\\
&=\mathbf{w}^t-\lambda\mathbf{g}^t-\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^t-\mathbf{a}^t\\
&=\hat{\mathbf{w}}^t-\lambda\mathbf{g}^t-\mathbf{a}^t.
\end{aligned}\tag{71}$$

Then, we complete the proof. ■

D. Proof of *Theorem 2*

Based on $\mathbf{u}_k^t=\lambda\mathbf{g}_k^t$ and $\mathbf{c}_{k,i}^t=\lambda\mathbf{g}_{k,i}^t-\mathcal{C}(\lambda\mathbf{g}_{k,i}^t)$, we have

$$\mathbf{w}^{t+1}=\mathbf{w}^t-\lambda\mathbf{g}^t+\frac{1}{D}\sum_{k=1}^{K}D_k\mathbf{c}_k^t-\mathbf{a}^t.\tag{72}$$

According to **Assumption 3**, in each communication round t, the global loss function $\mathcal{F}(\cdot)$ with $\mathbf{w}^{t+1}$ is bounded by

$$\mathcal{F}(\mathbf{w}^{t+1}) \leq \mathcal{F}(\mathbf{w}^t) + \left(\mathbf{w}^{t+1} - \mathbf{w}^t\right)^\top \nabla \mathcal{F}(\mathbf{w}^t) + \frac{L}{2}\left\|\mathbf{w}^{t+1} - \mathbf{w}^t\right\|_2^2. \tag{73}$$

By **Assumption 4**,

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{F}(\mathbf{w}^{t+1})\right] \\
&\leq \mathcal{F}(\mathbf{w}^t) + \mathbb{E}\left[\left(\mathbf{w}^{t+1} - \mathbf{w}^t\right)\right]^\top \nabla \mathcal{F}(\mathbf{w}^t) \\
&\quad + \frac{L}{2}\mathbb{E}\left\|\mathbf{w}^{t+1} - \mathbf{w}^t\right\|_2^2 \\
&= \mathcal{F}(\mathbf{w}^t) + (1-\lambda L)\,\mathbb{E}\left[\left(\frac{1}{D}\sum_{k=1}^{K} D_k \mathbf{c}_k^t - \mathbf{a}^t\right)\right]^\top \nabla \mathcal{F}(\mathbf{w}^t) \\
&\quad - \frac{\lambda}{2}(2-\lambda L)\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \frac{L}{2}\mathbb{E}\left\|\frac{1}{D}\sum_{k=1}^{K} D_k \mathbf{c}_k^t - \mathbf{a}^t\right\|_2^2 \\
&= \mathcal{F}(\mathbf{w}^t) - \frac{\lambda}{2}\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \frac{1}{2\lambda}\mathbb{E}\left\|\frac{1}{D}\sum_{k=1}^{K} D_k \mathbf{c}_k^t\right\|_2^2 \\
&\quad + \frac{1}{2\lambda}\mathbb{E}\left\|\mathbf{a}^t\right\|_2^2.
\end{aligned} \tag{74}$$

For a simpler expression without compromising generality, L is set to $1/\lambda$. Then, $\|\mathbf{c}_{k,i}^t\|_2^2 \leq (1 - s_i^t\delta)\,\lambda^2 G_i^2$ according to **Assumption 1**. Thus, by **Lemma 2** and Jensen's inequality, (74) is bounded as

$$\begin{aligned}
\mathbb{E}\left[\mathcal{F}(\mathbf{w}^{t+1})\right] &\leq \mathcal{F}(\mathbf{w}^t) - \frac{\lambda}{2}\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{2(b_i^t)^2 \lambda D^2} \\
&\quad + \frac{1}{2\lambda D}\sum_{k=1}^{K} D_k \mathbb{E}\left\|\mathbf{c}_k^t\right\|_2^2 \\
&\leq \mathcal{F}(\mathbf{w}^t) - \frac{\lambda}{2}\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{2(b_i^t)^2 \lambda D^2} \\
&\quad + \frac{1}{2\lambda D}\sum_{k=1}^{K} D_k \sum_{i=1}^{I}\mathbb{E}\left\|\mathbf{c}_{k,i}^t\right\|_2^2 \\
&\leq \mathcal{F}(\mathbf{w}^t) - \frac{\lambda}{2}\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 + \sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{2(b_i^t)^2 \lambda D^2} \\
&\quad + \frac{\lambda}{2}\sum_{i=1}^{I}\left(1 - s_i^t\delta\right) G_i^2.
\end{aligned} \tag{75}$$

According to **Assumption 2**, we rearrange the terms and take the average over t on both sides as

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\left\|\nabla \mathcal{F}(\mathbf{w}^t)\right\|_2^2 &\leq \frac{2}{\lambda}\left(\mathcal{F}(\mathbf{w}^1) - \mathcal{F}(\mathbf{w}^*)\right) \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}\frac{J_i s_i^t \sigma^2}{(b_i^t)^2 \lambda^2 D^2} \\
&\quad + \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{I}\left(1 - s_i^t\delta\right) G_i^2.
\end{aligned} \tag{76}$$

Then, we complete the proof. ∎

REFERENCES

[1] B. Liu, M. Ding, S. Shaham, W. Rahayu, F. Farokhi, and Z. Lin, "When machine learning meets privacy: A survey and outlook," *ACM Comput. Surv.*, vol. 54, no. 2, pp. 1–36, Mar. 2022.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Artif. Intell. Statist. (AISTATS)*, Fort Lauderdale, FL, USA, Apr. 2017, pp. 1273–1282.

[3] S. Niknam, H. S. Dhillon, and J. H. Reed, "Federated learning for wireless communications: Motivation, opportunities, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 46–51, Jun. 2020.

[4] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015, pp. 1–14.

[6] T. Dettmers, "8-bit approximations for parallelism in deep learning," in *Proc. ICLR*, San Juan, Puerto Rico, May 2016, pp. 1–14.

[7] D. Alistarh, D. Grubic, J. Li, R. Tomioka, and M. Vojnovic, "QSGD: Communication-efficient SGD via gradient quantization and encoding," in *Proc. Conf. Neural Inf. Process. Syst. (NeurIPS)*, Long Beach, CA, USA, Dec. 2017, pp. 1707–1718.

[8] J. Bernstein, Y.-X. Wang, K. Azizzadenesheli, and A. Anandkumar, "signSGD: Compressed optimisation for non-convex problems," in *Proc. Int. Conf. Mach. Learn. (ICML)*, Stockholm, Sweden, Jul. 2018, pp. 560–569.

[9] A. F. Aji and K. Heafield, "Sparse communication for distributed gradient descent," 2017, *arXiv:1704.05021*.

[10] N. Ivkin, D. Rothchild, E. Ullah, V. Braverman, I. Stoica, and R. Arora, "Communication-efficient distributed SGD with sketching," in *Proc. NeurIPS*, Vancouver, BC, Canada, Dec. 2019, pp. 13 142–13 152.

[11] Y. Jiang, S. Wang, V. Valls, B. J. Ko, W.-H. Lee, K. K. Leung, and L. Tassiulas, "Model pruning enables efficient federated learning on edge devices," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 12, pp. 10 374–10 386, Dec. 2023.

[12] D. Wen, K.-J. Jeon, and K. Huang, "Federated dropout—a simple approach for enabling federated learning on resource constrained devices," *IEEE Wireless Commun. Lett.*, vol. 11, no. 5, pp. 923–927, May 2022.

[13] N. Ström, "Scalable distributed DNN training using commodity GPU cloud computing," in *Proc. Int. Speech Commun. Assoc. (INTERSPEECH)*, Dresden, Germany, Sep. 2015, pp. 1488–1492.

[14] J. Jiang, F. Fu, T. Yang, and B. Cui, "SketchML: Accelerating distributed machine learning with data sketches," in *Proc. Int. Conf. Manag. Data (SIGMOD)*, Houston, TX, USA, Jun. 2018, pp. 1269–1284.

[15] D. Basu, D. Data, C. Karakus, and S. N. Diggavi, "Qsparse-local-SGD: Distributed SGD with quantization, sparsification, and local computations," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 217–226, May 2020.

[16] T. Vogels, S. P. Karimireddy, and M. Jaggi, "PowerSGD: Practical low-rank gradient compression for distributed optimization," in *Proc. NeurIPS*, Vancouver, BC, Canada, Dec. 2019, pp. 14 269–14 278.

[17] Y. Liu, W. Xu, G. Wu, Z. Tian, and Q. Ling, "Communication-censored ADMM for decentralized consensus optimization," *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2565–2579, May 2019.

[18] C. Li, G. Li, and P. K. Varshney, "Communication-efficient federated learning based on compressed sensing," *IEEE Internet Things J.*, vol. 8, no. 20, pp. 15 531–15 541, Oct. 2021.

[19] S. P. Karimireddy, Q. Rebjock, S. Stich, and M. Jaggi, "Error feedback fixes signSGD and other gradient compression schemes," in *Proc. ICML*, Long Beach, CA, USA, Jun. 2019, pp. 3252–3261.

[20] R. Kerkouche, G. Ács, C. Castelluccia, and P. Genevès, "Compression boosts differentially private federated learning," in *Proc. IEEE Eur. Symp. Security Privacy (EuroS&P)*, Vienna, Austria, Sep. 2021, pp. 304–318.

[21] Z. Tang, S. Shi, B. Li, and X. Chu, "GossipFL: A decentralized federated learning framework with sparsified and adaptive communication," *IEEE Trans. Parallel Distrib. Syst.*, vol. 34, no. 3, pp. 909–922, Mar. 2023.

[22] G. Zhu, Y. Wang, and K. Huang, "Broadband analog aggregation for low-latency federated edge learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 1, pp. 491–506, Jan. 2020.

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

[24] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Trans. Inf. Theory*, vol. 53, no. 10, pp. 3498–3516, Oct. 2007.

[25] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Trans. Signal Process.*, vol. 68, pp. 2155–2169, Mar. 2020.

[26] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.

[27] 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 Trans. Wireless Commun.*, vol. 20, no. 3, pp. 2120–2135, Mar. 2021.

[28] Y. Liu, D. Liu, G. Zhu, Q. Shi, and C. Zhong, "Over-the-air federated edge learning with error-feedback one-bit quantization and power control," 2023, *arXiv:2303.11319*.

[29] X. Fan, Y. Wang, Y. Huo, and Z. Tian, "1-bit compressive sensing for efficient federated learning over the air," *IEEE Trans. Wireless Commun.*, vol. 22, no. 3, pp. 2139–2155, Mar. 2023.

[30] Y. Sun, S. Zhou, Z. Niu, and D. Gündüz, "Time-correlated sparsification for efficient over-the-air model aggregation in wireless federated learning," in *Proc. IEEE Int. Conf. Commun. (ICC)*, Seoul, Korea, Republic of, May 2022, pp. 3388–3393.

[31] H. Ma, X. Yuan, Z. Ding, D. Fan, and J. Fang, "Over-the-air federated multi-task learning via model sparsification, random compression, and turbo compressed sensing," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4974–4988, Jul. 2023.

[32] Y. Sun, S. Zhou, and D. Gündüz, "Energy-aware analog aggregation for federated learning with redundant data," in *Proc. IEEE ICC*, Dublin, Ireland, Jun. 2020, pp. 1–7.

[33] X. Fan, Y. Wang, Y. Huo, and Z. Tian, "Joint optimization of communications and federated learning over the air," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4434–4449, Jun. 2022.

[34] W. Guo, R. Li, C. Huang, X. Qin, K. Shen, and W. Zhang, "Joint device selection and power control for wireless federated learning," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 8, pp. 2395–2410, Aug. 2022.

[35] N. Zhang and M. Tao, "Gradient statistics aware power control for over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 20, no. 8, pp. 5115–5128, Aug. 2021.

[36] X. Cao, G. Zhu, J. Xu, Z. Wang, and S. Cui, "Optimized power control design for over-the-air federated edge learning," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 342–358, Jan. 2022.

[37] X. Cao, G. Zhu, J. Xu, and S. Cui, "Transmission power control for over-the-air federated averaging at network edge," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 5, pp. 1571–1586, May 2022.

[38] G. Zhu, J. Xu, K. Huang, and S. Cui, "Over-the-air computing for wireless data aggregation in massive IoT," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 57–65, Aug. 2021.

[39] S. Zheng, Z. Huang, and J. T. Kwok, "Communication-efficient distributed blockwise momentum SGD with error-feedback," in *Proc. NeurIPS*, Vancouver, BC, Canada, Dec. 2019, pp. 11 450–11 460.

[40] E. Ozfatura, K. Ozfatura, and D. Gündüz, "Time-correlated sparsification for communication-efficient federated learning," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Melbourne, Australia, Jul. 2021, pp. 461–466.

[41] Y. You, J. Li, S. Reddi, J. Hseu, S. Kumar, S. Bhojanapalli, X. Song, J. Demmel, K. Keutzer, and C.-J. Hsieh, "Large batch optimization for deep learning: Training bert in 76 minutes," in *Proc. ICLR*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–37.

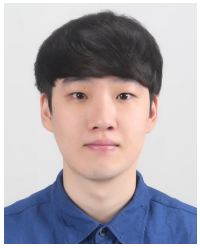
[42] S. Lee, T. Zhang, and A. S. Avestimehr, "Layer-wise adaptive model aggregation for scalable federated learning," in *Proc. AAAI Conf. Artif. Intell.*, Washington, DC, USA, Feb. 2023, pp. 8491–8499.

[43] A. Şahin and R. Yang, "A survey on over-the-air computation," *IEEE Commun. Surveys Tuts.*, vol. 25, no. 3, pp. 1877–1908, 3rd Quart. 2023.

[44] J. Liu, Z. Xu, R. Shi, R. C. C. Cheung, and H. K. H. So, "Dynamic sparse training: Find efficient sparse network from scratch with trainable masked layers," in *Proc. ICLR*, Addis Ababa, Ethiopia, Apr. 2020, pp. 1–14.

[45] Y. Lin, S. Han, H. Mao, Y. Wang, and W. J. Dally, "Deep gradient compression: Reducing the communication bandwidth for distributed training," in *Proc. ICLR*, Vancouver, BC, Canada, Apr. 2018, pp. 1–14.

Junsuk Oh received the B.S. and M.S. degrees in the School of Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2021 and 2023, respectively. He is currently pursuing a Ph.D. degree in the School of Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include wireless communication networks, distributed computing networks, signal processing and compression, over-the-air computation, and federated learning.



Donghyun Lee received his B.S. degree in Computer Science and Engineering from Dongguk University, South Korea, in 2020. He received his M.S. degree in Computer Science and Engineering from Chung-Ang University, Seoul, South Korea, in 2022. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include wireless networks, energy-efficient networks, wireless sensor networks, and multiple access.

Dongwook Won received his B.S. and M.S. degrees in Electronics and Communications Engineering from Kwangwoon University, Seoul, South Korea in 2020 and 2022, respectively. He is currently pursuing a Ph.D. degree in Computer Science and Engineering at Chung-Ang University, Seoul, South Korea. His research interests include semantic communications, 5G and 6G networks, and satellite communications.

Wonjong Noh received the B.S., M.S., and Ph.D. degrees from the Department of Electronics Engineering, Korea University, Seoul, Korea, in 1998, 2000, and 2005, respectively. From 2005 to 2007, he conducted his postdoctoral research at Purdue University, IN, USA, and the University of California at Irvine, CA, USA. From 2008 to 2014, he was a Principal Research Engineer with Samsung Advanced Institute of Technology, Samsung Electronics, Korea. He is currently an Professor at the School of Software at Hallym University, College of Information Science, Korea. His current research interests include fundamental capacity analysis and optimizations in 5G/6G wireless communication and networks, intelligent LEO satellite networks, federated mobile edge computing, machine learning-based systems control, and big-data based healthcare and medical system design. He received a government postdoctoral fellowship from the Ministry of Information and Communication, Korea, in 2005. He was also a recipient of the Samsung Best Paper Gold Award in 2010, the Samsung Patent Bronze Award in 2011, and the Samsung Technology Award in 2013. He has served numerous international conferences as a TPC member or an organizing committee member such as IEEE Globecom, WCNC, ICOIN, ICTC, ICUFN, ICMIC, JCCI and ICAIIC.

Sungrae Cho received the B.S. and M.S. degrees in Electronics Engineering from Korea University, Seoul, South Korea, in 1992 and 1994, respectively, and the Ph.D. degree in Electrical and Computer Engineering from the Georgia Institute of Technology, Atlanta, GA, USA, in 2002.

He is a Professor with the School of Computer Sciences and Engineering, Chung-Ang University (CAU), Seoul. Prior to joining CAU, he was an Assistant Professor with the Department of Computer Sciences, Georgia Southern University, Statesboro, GA, USA, from 2003 to 2006, and a Senior Member of Technical Staff with the Samsung Advanced Institute of Technology (SAIT), Kiheung, South Korea, in 2003. From 1994 to 1996, he was a Research Staff Member with the Electronics and Telecommunications Research Institute (ETRI), Daejeon, South Korea. From 2012 to 2013, he held a Visiting Professorship with the National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA. He has been a KICS fellow since 2021. He received numerous awards including Haedong Best Researcher of 2022 in Telecommunications and Award of Korean Ministry of Science and ICT in 2021.

His current research interests include wireless networking, network intelligence, and network optimization. He has been an Editor-in-Chief (EIC) of ICT Express (Elsevier) since 2024, a Subject Editor of IET Electronics Letter since 2018, an Executive Editor of Wiley Transactions on Emerging Telecommunications Technologies since 2023, and was an Area Editor of Ad Hoc Networks Journal (Elsevier) from 2012 to 2017. He has served numerous international conferences as a general chair, TPC chair, or an organizing committee chair, such as IEEE ICC, IEEE SECON, IEEE ICCE, ICOIN, ICTC, ICUFN, APCC, TridentCom, and the IEEE MASS.