MIMO Beamforming and Signal Modulation Design for Federated Learning Optimization

Nuocheng Yang*, Sihua Wang*, Mingzhe Chen†, Cong Shen‡, Changchuan Yin*, Christopher G. Brinton§

*Beijing Laboratory of Advanced Information Network, Beijing University of Posts and Telecommunications, Beijing, China, Emails: {YangNuoCheng,sihuawang,ccyin}@bupt.edu.cn.

†Department of Electrical and Computer Engineering and Institute for Data Science and Computing, University of Miami, Coral Gables, FL, 33146 USA, Email: mingzhe.chen@miami.edu.

‡Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA 22904 USA, Email: cong@virginia.edu.

§School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA, Email: cgb@purdue.edu.

***Abstract*—In this paper, we consider the optimization of federated learning (FL) over a realistic wireless multiple-input multiple-output (MIMO) communication system with digital modulation and over-the-air computation (AirComp). In such a system, MIMO devices transmit their locally trained FL models to a parameter server (PS) using beamforming to maximize the number of devices scheduled for transmission. AirComp enables efficient wireless model aggregation by the PS in bandwidth-limited settings. However, wireless channel fading can produce distortions in AirComp-based FL. To tackle this challenge, we develop a novel aggregation scheme that combines digital modulation with AirComp to mitigate wireless fading while ensuring communication efficiency. We formulate this as a joint transmit-receive beamforming design optimization problem which dynamically adjusts the beamforming matrices to minimize the FL training loss with transmission errors. To solve this problem based on limited information at the PS, we employ an artificial neural network (ANN) to estimate the local FL models of all devices. Then, we derive a closed-form optimal design of the transmit and receive beamforming matrices based on predicted FL models. Numerical evaluations validate the advantages of the proposed methodology in terms of model training performance compared with baselines.**

***Index Terms*—Federated learning, MIMO, AirComp, digital modulation.**

I. Introduction

In federated learning (FL) [1], a parameter server (PS) and edge devices iteratively exchange machine learning models over wireless channels. As a result, FL performance can be significantly affected by imperfect and dynamic wireless transmissions on both the uplink and downlink. Compared with the PS broadcasting the aggregated FL models to the edge devices, the edge devices uploading their locally trained models to the PS is more challenging due to their limited resource blocks [2].

To tackle this challenge, over-the-air computation (i.e., AirComp) techniques have recently been integrated into FL [3], [4]. Instead of decoding the individual local models of each device and then aggregating, AirComp allows edge devices to transmit their model parameters simultaneously over the same radio resources and decode the global average model directly at the PS. However, most of the existing works in this direction [5]–[10] have focused on the use of AirComp for analog modulation. Analog AirComp omits elegant FL convergence analysis results, but is not suitable for wireless communication systems that use digital modulations since these works do not consider source coding and channel coding. This motivates the investigation of AirComp-based FL over digital modulations in wireless systems.

A few recent works [11]–[14] have studied the implementation of AirComp FL over digital modulations. The authors in [11] designed one-bit quantization and modulation schemes for edge devices. In [12], the authors designed a joint channel decoding and aggregation decoding schemes based on binary phase shift keying (BPSK) modulation for AirComp FL. The authors in [13] evaluated the performance of FL gradient quantization in digital AirComp. The authors in [14] proposed a digital transmission protocol tailored to FL over wireless device-to-device networks for simultaneous transmissions. However, these prior works [11]–[14] mainly used low order digital modulation (i.e., BPSK) and hence their designed AirComp FL are not tailored to modern wireless systems that use high-order digital modulation schemes such as quadrature amplitude modulation (QAM). This is because the transmitted symbols that are processed by low-order digital modulation (such as the symbols -1 and +1 in BPSK) can be linearly superimposed. This linear superimposition poses challenges to high-order digital modulation schemes with complex mapping relationships between bits and symbols (such as Gray code).

In this paper, we develop a novel AirComp methodology for FL with high-order digital modulations in multiple-input multiple-output (MIMO) communication systems. In doing so, we make the following key contributions:

- We propose a novel AirComp-based MIMO system in which distributed wireless devices modulate their trained local FL parameters into symbols and transmit these modulated symbols to a PS that directly generates the global FL model via its received symbols.
- To optimize the FL training performance, the PS and devices must dynamically adjust the transmit and receive beamforming matrices over unreliable wireless channels. To this end, we formulate the joint transmit-receive beamforming matrix design problem as an optimization whose goal is to minimize FL training loss.
- In our problem, the optimal transmit and receive beam-

N. Yang, S. Wang and C. Yin were supported by the National Natural Science Foundation of China under Grants 61671086 and 61629101, and in part by Beijing Natural Science Foundation under Grant L223027 and 111 Project under Grant B17007. C. Brinton was supported in part by the U.S. National Science Foundation under grant CNS-2212565 and CNS-2146171.

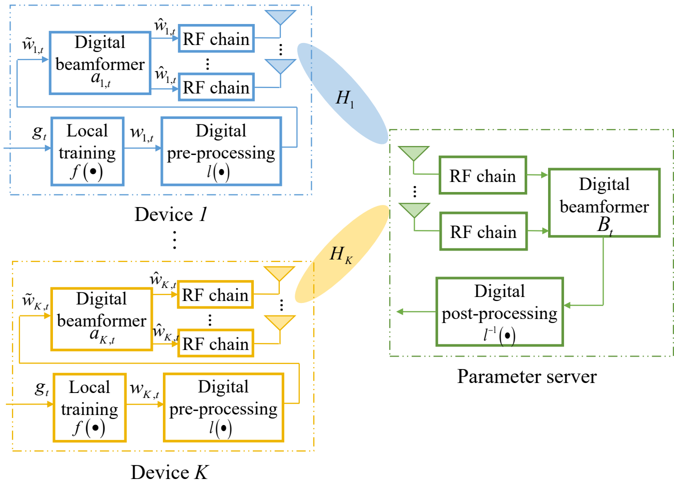

Fig. 1. The structure of FL deployed over multiple wireless devices and one PS in a MIMO communication system.

forming matrices depend on the channel environment and the trained FL model parameters, which are not known at the PS in advance. Hence, we employ an artificial neural network (ANN)-based algorithm to predict the FL model parameters of all devices. Given the predicted parameters, we analytically derive a closed-form solution for optimal design of the transmit and receive beamforming matrices based on the channel conditions.

- We conduct numerical evaluations on a real-world machine learning task dataset. Our results show that our methodology yields better model training performance characteristics than digital AirComp-based MIMO systems employing analog modulation and BPSK.

The rest of this paper is organized as follows. The system model and problem formulation for the AirComp-based MIMO FL system are described in Section II. Section III analyzes the convergence of the designed FL framework and derives a closed-form optimal design of the transmit and receive beamforming matrices based on the result. In Section IV, our numerical evaluation is presented and discussed. Finally, conclusions are drawn in Section V.

II. System Model and Problem Formulation

A. MIMO FL System

We consider a federated learning (FL) algorithm implemented over a wireless network, as shown in Fig. 1. K wireless edge devices train their individual machine learning models and send the machine learning parameters to a central PS through a noisy wireless MAC. In the considered model, the PS is equipped with N_r antennas while each device k is equipped with N_t antennas.

Each device has N_k training data samples and each training data sample n in device k consists of an input feature vector $\boldsymbol{x}_{k,n} \in \mathbb{R}^{N_I \times 1}$ and a corresponding label vector $\boldsymbol{y}_{k,n} \in \mathbb{R}^{N_O \times 1}$. The objective of training is to minimize the global loss function over all data samples, which is given by

$$F(\boldsymbol{g}) = \min_{\boldsymbol{g}} \frac{1}{N} \sum_{k=1}^{K} \sum_{n=1}^{N_k} f\left(\boldsymbol{g}, \boldsymbol{x}_{k,n}, \boldsymbol{y}_{k,n}\right), \tag{1}$$

where $\boldsymbol{g} \in \mathbb{R}^{V \times 1}$ is a vector that represents the global FL model of dimension V trained across the devices with $N = \sum_{k=1}^{K} N_k$ being the total number of training data samples of all devices. $f\left(\boldsymbol{g}, \boldsymbol{x}_{k,n}, \boldsymbol{y}_{k,n}\right)$ is a loss function of each device k.

To minimize the global loss function in (1) via a distributed manner, each device must update its FL model using its local dataset with stochastic gradient descent (SGD), which can be expressed as

$$\boldsymbol{w}_{k,t} = \boldsymbol{g}_t - \lambda \sum_{n \in \mathcal{N}_{k,t}} \frac{\partial f\left(\boldsymbol{g}, \boldsymbol{x}_{k,n}, \boldsymbol{y}_{k,n}\right)}{\partial \boldsymbol{g}}, \tag{2}$$

where λ is the learning rate, $\mathcal{N}_{k,t}$ is the subset of training data samples (i.e., minibatch) selected from device k's training dataset $\mathcal{N}_k$ at iteration t and $\boldsymbol{w}_{k,t}$ is the updated local FL model of device k at iteration t.

Given $\boldsymbol{w}_{k,t}$, distributed devices must simultaneously exchange their model parameters with the PS via bandwidth-limited wireless fading channels for model aggregation. The model aggregation is given by

$$\boldsymbol{g}_t = \sum_{k=1}^{K} \frac{|\mathcal{N}_k|}{N} \boldsymbol{w}_{k,t}, \tag{3}$$

where $|\mathcal{N}_k|$ represents the number of data samples in $\mathcal{N}_k$.

To ensure all devices can participate in FL model exchanging via wireless fading channels, each device adopts digital modulation to mitigate wireless fading and the PS adopts beamforming to maximize the number of devices scheduled for FL parameter transmission. Next, we will mathematically introduce the FL training and transmission process integrated with digital modulation in the considered MIMO communication system. In particular, we first introduce our designed digital modulation process that consists of two steps: (i) digital pre-processing at the devices and (ii) digital post-processing at the PS.

B. Digital Pre-Processing at the Devices

To transmit $\boldsymbol{w}_{k,t}$ over wireless fading channels, each device k leverages digital pre-processing to represent each numerical FL parameter in $\boldsymbol{w}_{k,t}$ using a symbol vector, which is

$$\hat{\boldsymbol{w}}_{k,t} = l\left(\boldsymbol{w}_{k,t}\right), \tag{4}$$

where $\hat{\boldsymbol{w}}_{k,t} \in \mathbb{R}^{L}$ is a modulated symbol vector with L being the number of symbols. $l\left(\cdot\right)$ is the digital pre-processing function that combines decimal-to-binary conversion and digital modulation where the decimal-to-binary conversion is used to represent each numerical FL parameter with a binary coded bit-interleaved vector and the digital modulation is used to modulate several binary bits as a symbol. We use rectangular M-quadrature-amplitude modulation (QAM) for digital modulation and it can be extended to other types of digital modulation schemes.

In our model, each device sends $\hat{\boldsymbol{w}}_{k,t}$ to the PS at each iteration t using full digital beamforming with low RF complexity. Given the transmit beamforming matrix $\boldsymbol{A}_{k,t} \in \mathbb{C}^{N_t \times L}$ and the maximal transmit power P_0 at device k, the power constraint can be expressed as

$$\mathbb{E}\left(\left|\boldsymbol{A}_{k,t}\hat{\boldsymbol{w}}_{k,t}\right|^2\right) = \left|\boldsymbol{A}_{k,t}\right|^2 \leq P_0. \tag{5}$$

C. Post-Processing at the PS

Considering the multiple access channel property of wireless communication, the received signal at the PS is given by

$$\boldsymbol{s}_t\left(\boldsymbol{A}_t\right) = \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{A}_{k,t} \hat{\boldsymbol{w}}_{k,t} + \boldsymbol{n}_t \tag{6}$$

where $\boldsymbol{A}_t = [\boldsymbol{A}_{1,t}, \cdots, \boldsymbol{A}_{K,t}]$ denotes the transmit beamforming matrices of all devices, $\boldsymbol{H}_k \in \mathbb{C}^{N_r \times N_t}$ denotes the MIMO channel vector for the link from device k to the PS, and $\boldsymbol{n}_t \in \mathbb{C}^{N_r}$ denotes additive white Gaussian noise. The entries of $\boldsymbol{H}_k$ and $\boldsymbol{n}_t$ are assumed to be independent and identically distributed (i.i.d.) complex Gaussian variables with zero mean.

Since $\boldsymbol{s}_t\left(\boldsymbol{A}_t\right)$ is the weighted sum of all users' local FL models, we consider directly generating the global FL model $\boldsymbol{g}_{t+1}$ from $\boldsymbol{s}_t(\boldsymbol{A}_t)$. This is a major difference between the existing works and this paper. The digital beamformer output signal can be expressed as

$$\hat{\boldsymbol{s}}_t\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right) = {\boldsymbol{B}_t}^{\mathrm{H}} \boldsymbol{s}_t(\boldsymbol{A}_t), \tag{7}$$

where $\boldsymbol{B}_t \in \mathbb{C}^{N_r \times L}$ is the digital receive beamforming matrix.

Given the received symbol vector $\hat{\boldsymbol{s}}_t\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right)$, the PS can reconstruct the numerical parameters in global FL model $\boldsymbol{g}_{t+1}$, which can be expressed as

$$\boldsymbol{g}_{t+1}\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right) = l^{-1}\left(\hat{\boldsymbol{s}}_t\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right)\right), \tag{8}$$

where $l^{-1}(\cdot)$ is the inverse function with respect to $l(\cdot)$ that combines the binary-to-decimal function and the digital demodulation function.

D. Problem Formulation

Next, we introduce our optimization problem. Our goal is to minimize the FL training loss by designing the transmit and receive beamforming matrices under the total transmit power constraint of each device, which is formulated as follows:

$$\min_{\boldsymbol{B}, \boldsymbol{A}} \; F\left(\boldsymbol{g}\left(\boldsymbol{B}_T, \boldsymbol{A}_T\right)\right), \tag{9}$$

$$\text{s.t.} \;\; \left|\boldsymbol{A}_{k,t}\right|^2 \leq P_0, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}. \tag{9a}$$

where $\boldsymbol{A} = [\boldsymbol{A}_1, \ldots, \boldsymbol{A}_T]$ and $\boldsymbol{B} = [\boldsymbol{B}_1, \ldots, \boldsymbol{B}_T]$ are the transmit and receive beamforming matrices for all iterations, respectively. T is a constant which is large enough to guarantee the convergence of FL.

From (9), we can see that the FL training loss $F\left(\boldsymbol{g}\left(\boldsymbol{B}_T, \boldsymbol{A}_T\right)\right)$ depends on the global FL model $g\left(\boldsymbol{B}_T, \boldsymbol{A}_T\right)$ that is trained iteratively. Meanwhile, as shown in (6) and (7), edge devices and the PS must dynamically adjust $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ based on current FL model parameters to minimize the gradient deviation caused by AirComp in the considered MIMO system with digital modulation.

III. Beamforming Optimization Algorithms

In this section, we turn to solving (9). Our methodology consists of two parts: (i) gradient vector forecasting and (ii) transmit-receive beamforming. In the second part, we analyze the convergence of FL so as to find the relationship between digital beamforming matrices $\boldsymbol{A}_t$, $\boldsymbol{B}_t$, predicted FL parameters, and FL training loss in (9). This leads us to a closed-form optimal design of $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$.

A. Gradient Vector Forecasting

A key challenge in solving (9) is that the PS does not know the gradient vector of each edge device in advance. Hence, the PS cannot proactively adjust the receive beamforming matrix using traditional optimization algorithms. To tackle this, we employ neural networks to predict the local FL model update of each device, since the optimal $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ depend on the trained FL parameters as shown in (6) and (8).

Finding a relationship among each device's local FL model updates at different iterations is a regression task. Fully-connected multilayer perceptrons (MLPs) in ANNs are known to excel at this. Thus, we choose to use MLPs instead of other neural networks such as recurrent neural networks (RNNs). The adopted MLP can capture the relationship between the input vector $\boldsymbol{g}_{k,t-1}$ and the output vector $\Delta\widetilde{\boldsymbol{w}}_{k,t}$ that is the predicted FL model update via a single hidden layer trained by an online gradient descent method.

B. Transmit-Receive Beamforming

Having the predicted local FL model updates $\Delta\widetilde{\boldsymbol{w}}_{k,t}$, the PS can optimize the beamforming matrices $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$ to minimize the FL training loss defined in (9). To this end, the PS must find the tightest gap between the optimal FL model and the current trained FL model for choosing optimal $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$. The update process of FL model at each iteration t can be given by the following lemma.

Lemma 1. Given the optimal global FL model $\boldsymbol{g}^*$, the current global FL model $\boldsymbol{g}_t$, the transmit beamforming matrix $\boldsymbol{A}_t$, and the receive beamforming matrix $\boldsymbol{B}_t$, the upper bound of $\mathbb{E}\left(F\left(\boldsymbol{g}_{t+1}(\boldsymbol{A}_t, \boldsymbol{B}_t)\right) - F\left(\boldsymbol{g}^*\right)\right)$ can be given by [15]

$$\begin{aligned}\mathbb{E}\left(F\left(\boldsymbol{g}_{t+1}(\boldsymbol{A}_t, \boldsymbol{B}_t)\right) - F\left(\boldsymbol{g}^*\right)\right) &\leqslant \mathbb{E}\left(F\left(\boldsymbol{g}_t\right) - F\left(\boldsymbol{g}^*\right)\right) \\ - \frac{1}{2L}\left\|\nabla F\left(\boldsymbol{g}_t\right)\right\|^2 &+ \frac{1}{2L}\mathbb{E}\left(\left\|\boldsymbol{e}_t\right\| + \left\|\hat{e_t}(\boldsymbol{A}_t, \boldsymbol{B}_t)\right\|\right)^2, \end{aligned} \tag{10}$$

where

$$\boldsymbol{e}_t = \frac{\sum_{k=1}^{K} l\left(\boldsymbol{w}_{k,t}\right)}{\sum_{k=1}^{K} \left|\mathcal{N}_{k,t}\right|} - l^{-1}\left(\frac{\sum_{k=1}^{K} l\left(\Delta\widetilde{\boldsymbol{w}}_{k,t}\right)}{\sum_{k=1}^{K} \left|\mathcal{N}_{k,t}\right|}\right) \tag{11}$$

with the first term being the gradient trained by selected data samples and the second term being the gradient predicted by the adopted MPL and

$$\hat{\boldsymbol{e}}_t(\boldsymbol{A}_t, \boldsymbol{B}_t) = l^{-1}\left(\frac{\sum_{k=1}^{K} l\left(\Delta\widetilde{\boldsymbol{w}}_{k,t}\right)}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right) - l^{-1}\left(\frac{\hat{\boldsymbol{s}}_t\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right)}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right). \tag{12}$$

From Lemma 1, we can see that since $\boldsymbol{e}_t$ does not depend on $\boldsymbol{A}_t$ or $\boldsymbol{B}_t$, the optimization of the digital beamforming matrices cannot minimize $\boldsymbol{e}_t$. In consequence, we can only minimize $\|\hat{\boldsymbol{e}}_t\|$ to decrease the gap between the FL training loss at iteration $t+1$ and the optimal FL training loss (i.e., $\mathbb{E}\left(F\left(\boldsymbol{g}_{t+1}\right) - F\left(\boldsymbol{g}^*\right)\right)$). Thus, problem (9) can be rewritten as

$$\min_{\boldsymbol{B}_t, \boldsymbol{A}_t} \left\| l^{-1}\left(\frac{\sum_{k=1}^{K} l\left(\Delta\widetilde{\boldsymbol{w}}_{k,t}\right)}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right) - l^{-1}\left(\frac{\hat{\boldsymbol{s}}_t\left(\boldsymbol{B}_t, \boldsymbol{A}_t\right)}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right)\right\|^2 \tag{13}$$

$$\text{s.t. } \left|\boldsymbol{A}_{k,t}\right|^2 \leq P_0, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}. \tag{13a}$$

In (13), the existence of the inverse function $l^{-1}(\cdot)$ defined in (8) significantly increases the complexity for finding optimal $\boldsymbol{A}_t$ and $\boldsymbol{B}_t$. Considering $l^{-1}(\cdot)$ that is used to demodulate the symbols into numerical FL parameters, the minimization of (13) is equivalent to minimizing the distance between $\frac{\sum_{k=1}^{K} l(\Delta\widetilde{\boldsymbol{w}}_{k,t})}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}$ that can be regarded as a constant and $\frac{\hat{\boldsymbol{s}}_t(\boldsymbol{B}_t, \boldsymbol{A}_t)}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}$ in the decision region of digital demodulation. Hence, in this section, we first derive the position of $\frac{\sum_{k=1}^{K} l(\Delta\widetilde{\boldsymbol{w}}_{k,t})}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}$ in the decision region and remove $l^{-1}(\cdot)$ from (13) for simplification. Then, we show a closed-form optimal design of the transmit and receive beamforming matrices.

Given $\Delta\widetilde{\boldsymbol{w}}_{k,t}$ and the digital pre-processing function $l(\cdot)$ defined in (4), the modulated symbol vector $\Delta\hat{\boldsymbol{w}}_{k,t} = l\left(\Delta\widetilde{\boldsymbol{w}}_{k,t}\right) = [\Delta\hat{w}_{k,t,1}^{\mathrm{I}}\Delta\hat{w}_{k,t,1}^{\mathrm{Q}}, \ldots, \Delta\hat{w}_{k,t,L}^{\mathrm{I}}\Delta\hat{w}_{k,t,L}^{\mathrm{Q}}]$ can be obtained where $\Delta\hat{w}_{k,t,i}^{\mathrm{I}}$ and $\Delta\hat{w}_{k,t,i}^{\mathrm{Q}}$ are the i-th in-phase symbol and quadrature symbol modulated by $\Delta\widetilde{\boldsymbol{w}}_{k,t}$, respectively. Since in-phase and quadrature-phase symbols that have vertical and horizontal decision regions are mutually independent, the value of $l^{-1}\left(\frac{\sum_{k=1}^{K}\Delta\hat{\boldsymbol{w}}_{k,t}}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right)$ can be obtained via individually analyzing the decision region of each in-phase and quadrature-phase symbols which are

$$\left|\frac{1}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{I}} - a_i^{\mathrm{I}}\right| \leqslant \frac{\xi}{2},$$
$$\left|\frac{1}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{Q}} - a_i^{\mathrm{Q}}\right| \leqslant \frac{\xi}{2}, \tag{14}$$

where $a_i^{\mathrm{I}}, a_i^{\mathrm{Q}} \in \mathcal{M} = \left\{\frac{1-\sqrt{M}}{2}\xi, \frac{3-\sqrt{M}}{2}\xi, \ldots, \frac{\sqrt{M}-1}{2}\xi\right\}$ are the constellation points in the decision region with $\mathcal{M}$ being the set of all constellation points. $\xi = \sqrt{\frac{4P_0}{\left(\sqrt{M}-1\right)^2}}$ is the minimum Euclidean distance between two constellation points. Using (14), a_i^{I} and a_i^{Q} are given by

$$a_{t,i}^{\mathrm{I}} = \left\{x \in \mathcal{M} : -\frac{\xi}{2} + \frac{\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{I}}}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|} \leqslant x \leqslant \frac{\xi}{2} + \frac{\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{I}}}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right\} \tag{15}$$

and

$$a_{t,i}^{\mathrm{Q}} = \left\{x \in \mathcal{M} : -\frac{\xi}{2} + \frac{\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{Q}}}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|} \leqslant x \leqslant \frac{\xi}{2} + \frac{\sum_{k=1}^{K}\Delta\hat{w}_{k,t,i}^{\mathrm{Q}}}{\sum_{k=1}^{K}|\mathcal{N}_{k,t}|}\right\}. \tag{16}$$

Given $\boldsymbol{a}_t^* = [a_{t,1}^{\mathrm{I}}a_{t,1}^{\mathrm{Q}}, \ldots, a_{t,L}^{\mathrm{I}}a_{t,L}^{\mathrm{Q}}]$, problem (13) can be rewritten as

$$\min_{\boldsymbol{B}_t, \boldsymbol{A}_t} \left\|\boldsymbol{a}_t^* - \boldsymbol{B}_t^{\mathrm{H}}\sum_{k=1}^{K}\boldsymbol{H}_k\boldsymbol{A}_{k,t}\Delta\hat{\boldsymbol{w}}_{k,t} - \boldsymbol{B}_t^{\mathrm{H}}\boldsymbol{n}_t\right\|^2 \tag{17}$$

$$\text{s.t. } \left\|\boldsymbol{A}_{k,t}\right\|^2 \leq P_0, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}. \tag{17a}$$

where $\Delta\hat{\boldsymbol{w}}_{k,t} = l\left(\Delta\widetilde{\boldsymbol{w}}_{k,t}\right)$ is a modulated symbol vector of $\Delta\widetilde{\boldsymbol{w}}_{k,t}$. To solve (17), we present an iterative algorithm that first optimizes transmit beamforming matrix $\boldsymbol{A}_t$ with fixed receive beamforming matrix $\boldsymbol{B}_t$, and then finds the optimal receive beamforming matrix $\boldsymbol{B}_t$ with optimized transmit beamforming matrix $\boldsymbol{A}_t$.

1) Optimization of Transmit Beamforming Matrix With Fixed Receive Beamforming Matrix: Given $\boldsymbol{B}_t$, problem (17) is a convex problem due to its convex objective function and

Algorithm 1 Proposed FL Over Digital AirComp-based MIMO System

Init: Global FL model $\boldsymbol{g}_0$, MIMO channel matrix $\boldsymbol{H}$.
for iterations $i = 0, 1, \cdots, T$ **do**
  for $k \in \{1, 2, \cdots, K\}$ in parallel over K devices **do**
    Each device calculates and returns $\boldsymbol{w}_{k,t}$ based on local dataset and $\boldsymbol{g}_t$ in (2).
    Each device leverages digital pre-processing to modulate each model parameter into a symbol.
    Each device sends the symbol vector $\hat{\boldsymbol{w}}_{n,k}$ to the PS using the optimized transmit beamforming matrix $\boldsymbol{A}_{k,t}$.
  end for
  The PS directly demodulates the global FL model $\boldsymbol{g}_{t+1}$ from the received superpositioned signal using (8).
  The PS predicts the local FL model $\hat{\boldsymbol{w}}_{k,t+1}$ of each device based on demodulated $\boldsymbol{g}_{t+1}$ using trained ANN.
  The PS proactively adjusts the transmit and receive beamforming matrices using the augmented Lagrangian method and broadcast the transmit beamforming matrix $\boldsymbol{A}_{k,t+1}$ to each device k.
end for

constraints, which can be optimally solved by using the dual method. The Lagrangian function of problem (17) will be

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{A}_t, \boldsymbol{\lambda}_t, \boldsymbol{\mu}_t\right) = & \left\| \boldsymbol{a}_t^* - {\boldsymbol{B}_t}^{\mathrm{H}} \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{A}_{k,t} \Delta \hat{\boldsymbol{w}}_{k,t} - {\boldsymbol{B}_t}^{\mathrm{H}} \boldsymbol{n}_t \right\|^2 \\ & - \sum_{k=1}^{K} \lambda_{k,t} \left(\left\| \boldsymbol{A}_{k,t} \right\|^2 - P_0 + \mu_{k,t} \right) \\ & + \frac{\alpha}{2} \left(\left\| \boldsymbol{A}_{k,t} \right\|^2 - P_0 + \mu_{k,t} \right)^2, \end{aligned} \tag{18}$$

where $\boldsymbol{\lambda}_t = [\lambda_{1,t}, \ldots, \lambda_{K,t}]$ and $\boldsymbol{\mu}_t = [\mu_{1,t}, \ldots, \mu_{K,t}]$ are the Lagrangian multiplier vectors, and α is the penalty parameter. Problem (18) can be solved by a traditional augmented Lagrangian method that approaches the optimal solution via alternating updating. However, finding $\boldsymbol{A}_{k,t}^*$ in (18) is complex since the dimension of $\boldsymbol{A}_{k,t}$ can be extremely large as the number of transmit antennas and transmitted symbols on each device increases. Thus, an inexact alternating direction method (IADM) is used to approach $\boldsymbol{A}_{k,t}^*$, $\boldsymbol{\lambda}_t^*$, and $\boldsymbol{\mu}_t^*$. Using the Karush Kuhn-Tucker (KKT) conditions for the optimal solution $\boldsymbol{A}_{k,t}^*$, the update rules for approaching $\boldsymbol{A}_{k,t}$, $\boldsymbol{\lambda}_t^*$, and $\boldsymbol{\mu}_t^*$ can be given by

$$\begin{aligned}\boldsymbol{A}_{k,t}^{i+1} = & \boldsymbol{A}_{k,t}^{i} - 2\tau \boldsymbol{A}_{k,t} \left(-\lambda_{k,t} + \alpha \left(\boldsymbol{A}_{k,t} - P_0 + \mu_{k,t} \right) \right) \\ & - \tau \boldsymbol{H}_t {\boldsymbol{B}_k}^{\mathrm{H}} \Delta \hat{\boldsymbol{w}}_{k,t} \left(\boldsymbol{a}_t^* - {\boldsymbol{B}_t}^{\mathrm{H}} \sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{A}_{k,t} \Delta \hat{\boldsymbol{w}}_{k,t} - {\boldsymbol{B}_t}^{\mathrm{H}} \boldsymbol{n}_t \right), \end{aligned} \tag{19}$$

$$\lambda_{k,t}^{i+1} = \lambda_{k,t}^{i} - \alpha \left(\left\| \boldsymbol{A}_{k,t}^{i+1} \right\|^2 - P_0 + \mu_{k,t}^{i+1} \right), \tag{20}$$

TABLE I
SIMULATION PARAMETERS

Parameters	Values	Parameters	Values
K	5	M	64
N_t	2	N_r	4
P_0	0.0001 W	ϵ	0.0001

$$\mu_{k,t}^{i+1} = \frac{\lambda_{k,t}^{i+1}}{\alpha} - \left\| \boldsymbol{A}_{k,t}^{i+1} \right\|^2 + P_0. \tag{21}$$

where τ is a dynamic step size and (19)-(21) will be repeated multiple times until $||\boldsymbol{A}_{k,t}^{i+1} - \boldsymbol{A}_{k,t}^{i}||^2 \leqslant \epsilon$ with ϵ being the tolerance error.

2) Optimization of Receive Beamforming Matrix With Fixed Transmit Beamforming Matrix: Once the transmit beamforming matrix $\boldsymbol{A}_t$ is determined, $\hat{\boldsymbol{e}}_t$ only depends on the rceive beamforming matrix $\boldsymbol{B}_t$. Since constraint (17a) is only determined by $\boldsymbol{A}_t$, the optimization problem in (17) with optimized $\boldsymbol{A}_t^*$ can be formulated as

$$\min_{\boldsymbol{B}_t} \left\| \boldsymbol{a}_t^* - {\boldsymbol{B}_t}^{\mathrm{H}} \left(\sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{A}_{k,t}^* \Delta \hat{\boldsymbol{w}}_{k,t} + \boldsymbol{n}_t \right) \right\|^2. \tag{22}$$

Given (22), the optimized $\boldsymbol{B}_t$ at time t is

$$\boldsymbol{B}_t^* = \left(\frac{\boldsymbol{a}_t^*}{\sum_{k=1}^{K} \boldsymbol{H}_k \boldsymbol{A}_{k,t}^* \Delta \hat{\boldsymbol{w}}_{k,t}} \right)^{\mathrm{H}}.$$

IV. NUMERICAL EVALUATION

For our simulations, we consider a circular network area having a radius $r = 1500$ m with one PS at its center serving $K = 5$ uniformly distributed devices. The other parameters used in simulations are listed in Table I. For comparison purposes, we use two baselines: a) an FL algorithm that is deployed in digital AirComp-based MIMO system with analog modulation [10] and b) an FL algorithm that is deployed in digital AirComp-based MIMO system with low order digital modulation [12]. For the ML task, we consider image classification on the standard MNIST dataset.

In Fig. 2, we show how the identification accuracy of all considered algorithms changes as the number of iterations varies when $\boldsymbol{n}_t$ = -50 dBW. Compared with the digital AirComp-based MIMO system with analog modulation and BPSK, the proposed digital AirComp method with 64 QAM can achieve up to 1% and 20% gains in terms of identification accuracy. The 1% gain stems from the fact that the digital modulation can better combat channel impairments and misalignments, thereby achieving accurate model aggregation even in unsatisfactory wireless channel conditions. The 20% gain stems from the fact that, each parameter must be quantized into -1 or $+1$ for transmission, thus resulting a quantization error and encumbering the FL performance in terms of identification accuracy.

Fig. 3 shows how the identification accuracy changes as the noise power increases. From this figure, we can see that, the identification accuracy decreases slightly as noise power

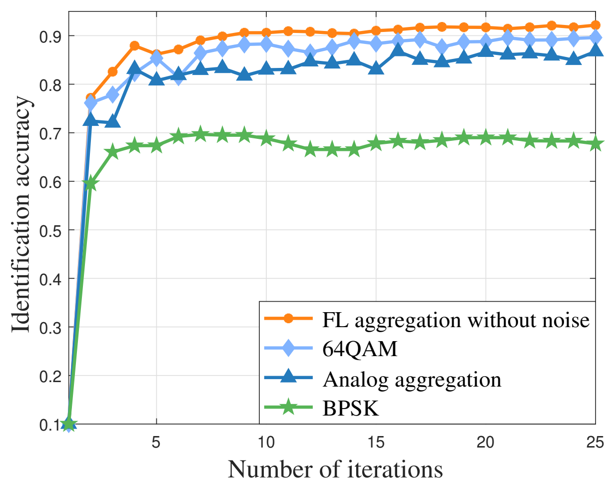

Fig. 2. Identification accuracy vs. number of iterations.

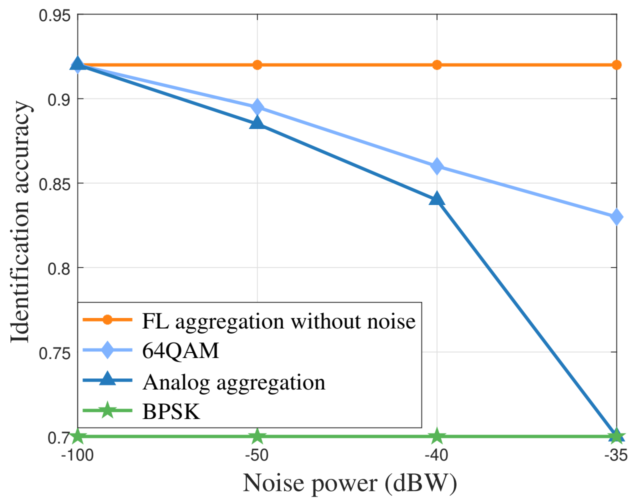

Fig. 3. Identification accuracy vs. noise power.

increases. This is because as noise power increases, the accuracy for FL parameter transmission decreases, which results in additional error in model aggregation process and decreases the identification accuracy. Fig. 3 also shows that the proposed digital AirComp method achieves the identification accuracy by up to 13% compared to the digital AirComp-based MIMO system with analog modulation and BPSK when noise power is -35dBW. This is due to the fact that the proposed method can eliminate the transmission error introduced by unsatisfactory channel environment via digital demodulation and generate an accuract FL model in model aggregation process. The identification accuracy of the digital AirComp-based MIMO system with BPSK remains a constant as the noise power increases. This is because that each parameter is quantized into -1 or $+1$ for transmission in BPSK, which maximizes decision thresholds in digital demodulation under same transmit power, thus resulting in an unchanged identification accuracy with low noise power. However, such digital demodulation method would introduce additional quantization error in FL which significantly affects the identification accuracy.

V. Conclusion

In this paper, we have developed a novel framework that enables the implementation of FL algorithms over a digital AirComp-based MIMO system. We have formulated an optimization problem that jointly considers transmit and receive beamforming matrices for the minimization of FL training loss. To solve this problem, we have analyzed the expected improvement of FL training loss between two adjacent iterations that depends on the digital modulation level, the number of devices, and the design of beamforming matrices. To find the tightest bound, we introduced an ANN based algorithm to estimate the local FL models of all devices and then, the optimal solution of beamforming matrices is determined based on the predicted FL model and the derived expected improvement of FL training loss. Numerical evaluation on real-world machine learning tasks demonstrated that the proposed method yields significant gains in classification accuracy and convergence speed compared with conventional approaches.

References

[1] Y. Oh, T.S. Jeon, M. Chen, and W. Saad, "Vector quantized compressed sensing for communication-efficient federated learning," in *2022 IEEE Globecom Workshops (GC Wkshps)*, Jan. 2022, pp. 365–370.

[2] M. Chen, D. Gündüz, K. Huang, W. Saad, M. Bennis, A. V. Feljan, and H. V. Poor, "Distributed learning in wireless networks: Recent progress and future challenges," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3579–3605, Oct. 2021.

[3] T. Sery, N. Shlezinger, K. Cohen, and Y. C. Eldar, "Over-the-air federated learning from heterogeneous data," *IEEE Transactions on Signal Processing*, vol. 69, pp. 3796–3811, June 2021.

[4] N. Zhang and M. Tao, "Gradient statistics aware power control for over-the-air federated learning," *IEEE Transactions on Wireless Communications*, vol. 20, no. 8, pp. 5115–5128, March 2021.

[5] X. Ma, H. Sun, Q. Wang, and R. Q. Hu, "User scheduling for federated learning through over-the-air computation," in *Proc. of IEEE Vehicular Technology Conference (VTC2021-Fall)*, Norman, OK, USA, Sept. 2021.

[6] X. Zhai, X. Chen, J. Xu, and D. W. Kwan Ng, "Hybrid beamforming for massive mimo over-the-air computation," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2737–2751, Jan. 2021.

[7] S. Xia, J. Zhu, Y. Yang, Y. Zhou, Y. Shi, and W. Chen, "Fast convergence algorithm for analog federated learning," in *Proc. of IEEE International Conference on Communications (ICC)*, Montreal, QC, Canada, June 2021.

[8] C. Xu, S. Liu, Z. Yang, Y. Huang, and K. K. Wong, "Learning rate optimization for federated learning exploiting over-the-air computation," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3742–3756, Dec. 2021.

[9] Y. Hu, M. Chen, M. Chen, Z. Yang, M. Shikh-Bahaei, H. V. Poor, and S. Cui, "Energy minimization for federated learning with IRS-assisted over-the-air computation," in *Proc. of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Oronto, ON, Canada, June 2021.

[10] S. Xia, J. Zhu, Y. Yang, Y. Zhou, Y. Shi, and W. Chen, "Fast convergence algorithm for analog federated learning," in *Proc. of IEEE International Conference on Communications (ICC)*, Montreal, QC, Canada, June 2021.

[11] 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, Nov. 2021.

[12] X. Zhao, L. You, R. Cao, Y. Shao, and L. Fu, "Broadband digital over-the-air computation for asynchronous federated edge learning," Available Online*: https://arxiv.org/abs/2111.10508*, Nov. 2021.

[13] M. M. Amiri and D. Gündüz, "Machine learning at the wireless edge: Distributed stochastic gradient descent over-the-air," *IEEE Transactions on Signal Processing*, vol. 68, pp. 2155–2169, March 2020.

[14] H. Xing, O. Simeone, and S. Bi, "Federated learning over wireless device-to-device networks: Algorithms and convergence analysis," *IEEE Journal on Selected Areas in Communications*, vol. 39, no. 12, pp. 3723–3741, Oct. 2021.

[15] M. Chen, Z. Yang, W. Saad, C. Yin, and H. V. Poorand S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 1, pp. 269–283, Oct. 2021.