



RIS-Assisted Federated Learning in Multi-Cell Wireless Networks

WANG Yiji, WEN Dingzhu, MAO Yijie, SHI Yuanming

(ShanghaiTech University, Shanghai 201210, China)

DOI: 10.12142/ZTECOM.202301004

<https://kns.cnki.net/kcms/detail/34.1294.TN.20230227.1850.002.html>,
published online February 28, 2023

Manuscript received: 2022-12-04

Abstract: Over-the-air computation (AirComp) based federated learning (FL) has been a promising technique for distilling artificial intelligence (AI) at the network edge. However, the performance of AirComp-based FL is decided by the device with the lowest channel gain due to the signal alignment property. More importantly, most existing work focuses on a single-cell scenario, where inter-cell interference is ignored. To overcome these shortages, a reconfigurable intelligent surface (RIS)-assisted AirComp-based FL system is proposed for multi-cell networks, where a RIS is used for enhancing the poor user signal caused by channel fading, especially for the device at the cell edge, and reducing inter-cell interference. The convergence of FL in the proposed system is first analyzed and the optimality gap for FL is derived. To minimize the optimality gap, we formulate a joint uplink and downlink optimization problem. The formulated problem is then divided into two separable nonconvex subproblems. Following the successive convex approximation (SCA) method, we first approximate the nonconvex term to a linear form, and then alternately optimize the beamforming vector and phase-shift matrix for each cell. Simulation results demonstrate the advantages of deploying a RIS in multi-cell networks and our proposed system significantly improves the performance of FL.

Keywords: federated learning (FL); reconfigurable intelligent surface (RIS); over-the-air computation (AirComp); multi-cell networks

Citation (IEEE Format): Y. J. Wang, D. Z. Wen, Y. J. Mao, et al., "RIS-assisted federated learning in multi-cell wireless networks," *ZTE Communications*, vol. 21, no. 1, pp. 25 - 37, Mar. 2022. doi: 10.12142/ZTECOM.202301004.

1 Introduction

With the development of the Internet of Things (IoT) and wireless technologies, recent years have witnessed an explosion of IoT devices and mobile data, which is of great significance for training AI models to enable various kinds of intelligent applications, such as auto-driving vehicles, equipment condition monitoring, and smart cities^[1-2]. However, conventional methods that upload massive distributed data to a cloud encounter huge communication overhead and violate data privacy. To overcome these problems, federated learning (FL) emerges as a promising solution, where a shared AI model is trained among multiple devices without raw data transmission^[3-6]. Specifically, there are three steps in each training iteration of FL. First, a central server generates an initial global model and then broadcasts the global model to the edge devices covered by it. Then, each edge device performs one or more steps of local training based on the received global model and local dataset to calculate a local model or gradient vector and uploads it to the central server. Finally, the central server aggregates all local information and updates the global model for the next communication round.

One main research direction of FL is to overcome the com-

munication bottleneck caused by frequent transmission of the high dimensional model and gradient vectors. To combat the influence of wireless communications, the authors in Ref. [7] proposed a joint learning and communication framework to minimize the FL loss function. Partial device participation approaches, such as random scheduling and proportional fairness, have been proposed for the rational allocation of limited communication resources in FL^[8]. To improve the communication efficiency of the FL uplink model aggregation, an over-the-air computation (AirComp) technique based on the waveform superposition characteristics of the multiple access channels (MACs) was proposed in Refs. [9 - 13], which realizes the summation calculation of the receiver function during information transmission. To overcome the bottleneck of limited communication bandwidth in the aggregation process, the authors in Ref. [14] presented a fast model aggregation method to improve the performance of FL by jointly optimizing beamforming vectors and device selection. In Ref. [15], a federated zeroth-order optimization (FedZO) algorithm based on AirComp was proposed to enable communication-efficient transmission by performing multiple local updates and partial device participation. Compared with the orthogonal multiple access (OMA) method, where the information of other users is regarded as interference, and the summation of all signals is

then calculated, i. e., computing after communication, AirComp greatly improves communication efficiency. The benefits of AirComp-based FL have motivated its application in the unmanned aerial vehicle (UAV)^[16-17] and reconfigurable intelligent surface (RIS)-enabled networks^[18-23].

The schemes mentioned above cannot solve the essential problem that wireless channel fading leads to poor signal strength of many devices, especially for AirComp-based FL, whose performance generally depends on the worst device in the network. To mitigate the effects of wireless channel fading, RIS is recognized as a revolutionary technology that achieves high spectrum and energy efficiency by reconfiguring the wireless channel environment at a low cost^[24-27]. The authors in Ref. [25] designed a RIS-assisted AirComp system to increase the performance of AirComp by optimizing the transceivers and RIS phase-shift. It was shown in Refs. [19-20] that configuring RISs in AirComp-based FL further reduced the error of model aggregation, thereby improving the learning performance. Considering the low latency and privacy-secure nature of FL, a differentially private FL system via RIS was proposed in Ref. [12] to achieve a better tradeoff between the learning performance and privacy under the constraints of privacy and power. In order to further reduce the aggregation error, a multi-RIS scenario was presented in Ref. [28], where both the base station and the user used one dedicated RIS to mitigate the effects of poor channels. However, all the aforementioned works are limited to a single-cell setting. In fact, considering a multi-cell scenario is more in line with practical large-scale network design^[29-31]. Due to the serious fading of the signal received by users at the cell edge, deploying RISs can relay the intended signal to enhance signal strengths for edge users and expand network coverage in multi-cell scenarios^[31-33]. Besides, the authors in Refs. [30] and [34] proved that deploying a RIS at the cell edge can achieve the highest performance gain compared with other RIS deployments. Most of the existing RIS-assisted multi-cell networks focus on communication-only system models, ignoring the application of FL. Although the multi-cell FL interference management was considered in Ref. [29], RIS was not considered to enhance the performance of FL. To the best of our knowledge, this is the first work that investigates AirComp-based FL in RIS-assisted multi-cell networks.

In this paper, we investigate a RIS-assisted AirComp-based FL system in multi-cell networks, where a RIS is deployed at the cell edge to help each cell complete different FL tasks. In the process of FL, we consider both the impact of downlink and uplink communications. For the fast aggregation of uplink gradients, we adopt AirComp to improve communication efficiency. However, the performance of AirComp-based FL is dependent on the device with the worst link gain (e.g., the cell-edge device with a large path loss). Besides, the inter-cell interference also degrades its performance. To address these is-

ues, we further deploy a RIS at the cell edge to enhance signal strength and mitigate inter-cell interference, thereby improving the FL performance. In our proposed system, there are some difficulties that we need to highlight. First, we consider both the impact of downlink model dissemination and that of uplink gradient aggregation, both are inevitably affected by channel fading, noise and inter-cell interference. It is different from most FL works, i. e., only uplink aggregation errors are considered. Second, considering the downlink influence makes the convergence analysis of our system more complicated. This derivation result is related to noise and inter-cell interference. Third, the optimization problems are non-convex and complex. We have to jointly optimize the beamforming vector and phase shift to improve the performance of our proposed system. The main contributions of this paper are summarized as follows:

- We propose a RIS-assisted AirComp-based FL system in two-cell networks, where a RIS is used for enhancing the signal of cell-edge devices during the process of both downlink and uplink transmission as well as for canceling the inter-cell interference. Then, we derive the convergence analysis of the proposed framework. The optimal gap of FL is determined by the uplink error and the downlink error of two cells, and each error contains channel fading, inter-cell interference and received noise.

- To maximize the learning performance for all cells, it is necessary to minimize the optimal gap. To this end, we decouple this optimization problem into two separate subproblems, respectively for the downlink and uplink optimization. Each subproblem requires a joint alternating optimization of beamforming vectors and phase-shift matrices. Since the optimization subproblems remain nonconvex, we first make a variable conversion and then utilize the successive convex approximation (SCA) method to approximate the problem. An alternative optimization algorithm is then proposed to solve each subproblem.

- Extensive simulations are performed to verify the performance of the proposed RIS-assisted FL system in two-cell networks. It shows that the proposed scheme can enhance the performance of the AirComp-based FL system by enhancing the signal strength and suppressing the inter-cell interference. In addition, the proposed algorithm guarantees fairness among cells.

The rest of this paper is organized as follows. Section 2 introduces the system model of RIS-assisted AirComp-based FL in a two-cell scenario. Section 3 provides the convergence analysis and the problem formulation. In Section 4, we propose an SCA-based joint alternating beamforming and phase-shift matrix optimization to minimize the upper bound of all cells. Simulation results are provided in Section 5 to support the advantages of the proposed system. Finally, we conclude this work in Section 6.

2 System Model

2.1 Network Model

As shown in Fig. 1, we mainly develop a RIS-assisted AirComp-based FL system in a two-cell network, where each cell has K single-antenna edge devices and one access point (AP), where each AP is equipped with N antennas. At the edge of two cells, we deploy a RIS to enhance the signal strength of edge devices, where the RIS has S passive reflecting elements. Edge device $k \in \mathcal{K}_l = \{1, 2, \dots, K\}$ is associated with AP $l \in \mathcal{L} = \{1, 2\}$ to complete information exchange under both downlink and uplink communications, where $\mathcal{K}_l \cap \mathcal{K}_j = \emptyset, \forall l \neq j$ and $l, j \in \mathcal{L}$. During the process of transmission, we assume that each AP knows the channel state information for all edge devices.

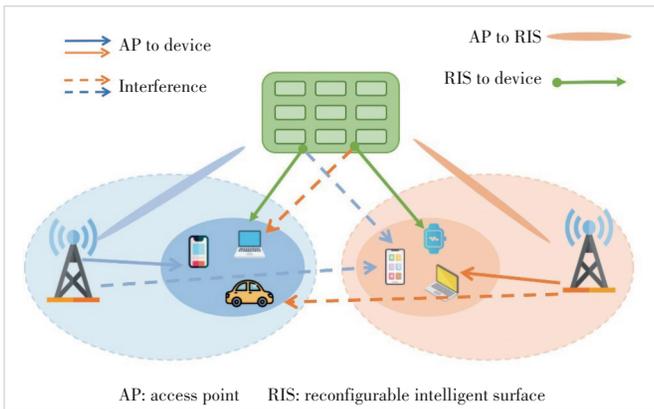
2.2 Federated Learning Model

In the two-cell FL system, each edge device $k \in \mathcal{K}_l$ has its own local dataset $\mathcal{D}_k = \{(x_i, y_i)\}_{i=1}^{D_k}$ with $D_k = |\mathcal{D}_k|$ data samples and each cell trains an individual FL model. The goal of FL is to collaboratively train a shared model $\mathbf{w}_l \in \mathbb{R}^d$ of dimension d without making any local dataset public. The local loss function for edge device $k \in \mathcal{K}_l$ is defined as

$$F_{l,k}(\mathbf{w}_l) = \frac{1}{D_k} \sum_{(x_i, y_i) \in \mathcal{D}_k} f_l(\mathbf{w}_l; x_i, y_i), \quad (1)$$

where $f_l(\cdot)$ is the sample-wise loss function defined by the learning task for cell l . In this work, we consider a general model where learning tasks of the two cells are different. Without loss of generality, all local datasets for users in the same cell are assumed to have the same size, i. e., $|\mathcal{D}_{k_1}| = |\mathcal{D}_{k_2}|, \forall k_1, k_2 \in \mathcal{K}_l$. As a result, the global loss function for the learning task in cell l can be expressed as

$$F_l(\mathbf{w}_l) = \frac{1}{K} \sum_{k \in \mathcal{K}_l} F_{l,k}(\mathbf{w}_l). \quad (2)$$



▲ Figure 1. RIS-assisted AirComp-based FL system in a two-cell network

Then the global model for the l -th cell is obtained by

$$\mathbf{w}_l^* = \arg \min_{\mathbf{w}_l \in \mathbb{R}^d} F_l(\mathbf{w}_l). \quad (3)$$

To achieve the above FL purpose, we utilize the federated stochastic gradient descent (FedSGD) algorithm to perform local updates, which means only part of the datasets participates in training. Specifically, at the t -th communication round, AP l and the edge devices perform the following three procedures:

1) Broadcasting: AP l broadcasts the current global model \mathbf{w}_l^t to the edge devices belonging to this cell l .

2) Local model update: Based on the received global model \mathbf{w}_l^t , each edge device $k \in \mathcal{K}_l$ performs a one-step local model update via the local mini-batch SGD algorithm, which is given by

$$\begin{aligned} \mathbf{w}_{l,k}^t &= \mathbf{w}_l^t - \frac{\zeta^t}{B} \sum_{(x_i, y_i) \in \mathcal{B}_k^t} \nabla f_l(\mathbf{w}_l^t; x_i, y_i, \mathcal{B}_k^t) = \\ &= \mathbf{w}_l^t - \zeta^t \nabla F_{l,k}(\mathbf{w}_l^t), \end{aligned} \quad (4)$$

where ζ^t denotes the learning rate and \mathcal{B}_k^t is the mini-batch dataset with size B_k^t . Besides, we let $\mathbf{p}_k^t = \nabla F_{l,k}(\mathbf{w}_l^t)$ denote the trained gradient information. Then all edge devices upload the computed gradient information \mathbf{p}_k^t .

3) Model aggregation and update: The AP aggregates the received local gradient information and then generates a new global model as:

$$\mathbf{w}_l^{t+1} = \mathbf{w}_l^t - \frac{\zeta^t}{K_l} \sum_{k \in \mathcal{K}_l} \mathbf{p}_k^t. \quad (5)$$

Algorithm 1 summarizes the above steps of FedSGD.

Algorithm 1: FedSGD

Input: Initialize the global model \mathbf{w}^0 , communication round T , local iteration epoch E , mini-batch dataset \mathcal{B} , and learning rate ζ .

for communication round $t = 1, 2, \dots, T$ **do**

AP broadcasts the global model \mathbf{w}^t to the edge devices;

Edge devices initial local model $\mathbf{w}_k^{t,0} = \mathbf{w}^t$ and make E local training;

for local iteration epoch $e = 1, 2, \dots, E$ **do**

$$\mathbf{w}_k^{t,e} \leftarrow \text{LocalSGD}(\mathbf{w}_k^{t,e-1}, \mathcal{B}_k^t)$$

end

Compute the cumulative gradient information $\nabla f_{l,k} \leftarrow \mathbf{w}_k^{t,e} - \mathbf{w}_k^{t,0}$;

Upload the gradient information and update the global model

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \zeta^t \nabla F_{l,k}(\mathbf{w}_l^t);$$

end

In the proposed two-cell system, we assume that these steps are synchronous in both cells and their gradient information is

uploaded to the AP. The synchronization can be enabled by AirShare^[35], which transmits the clock over the air and provides a distributed protocol. In the next section, we elaborate on the communication process of the proposed system following the procedure of FL.

2.3 Downlink Communication for RIS-Assisted FL System

From the perspective of communication, we utilize the universal frequency reuse technique to improve spectral efficiency. In other words, the two cells share the same frequency during both downlink and uplink communications, inevitably causing inter-cell interference.

Considering a round of downlink communications in cell l , AP l shares the global model with each edge device in cell l . However, in most of the existing works on FL, the process of broadcast is error-free, which indicates the edge device $k \in \mathcal{K}_l$ can accurately receive signals from AP l . In this subsection, we consider the effects of noise and inter-cell interference in downlink communications. Here, we omit the time index and denote the downlink transmitted signal from AP l to the edge device k as w_l . In addition, we assume w_l follows the standard Gaussian distribution, i.e., $w_l \sim \mathcal{CN}(0,1)$. However, the transmitted signals may go through poor channel conditions in the communication process, which results in a larger receive error at edge device k . To lift the accuracy of the received signal, we deploy a RIS to mitigate the distortion of signals.

Specifically, we let $\boldsymbol{\theta}_d = [\beta e^{j\theta_s^d}, \dots, \beta e^{j\theta_s^d}]$ represent the diagonal phase-shift matrix of the RIS in the downlink communication and $\boldsymbol{\Theta}_d = \text{diag}(\boldsymbol{\theta}_d)$ with $\theta_s^d \in [0, 2\pi]$ and $\beta \in [0, 1]$ is the amplitude reflection coefficient on the incident signal. To be specific, we set $\beta = 1$ in this paper and mainly consider the first reflected signal^[24], because the signal reflected by multiple times appears insignificant due to propagation loss. Subsequently, we let $\mathbf{h}_{l,k}^l \in \mathbb{C}^N$, $\mathbf{h}_{l,k}^r \in \mathbb{C}^S$, and $\mathbf{G}_l \in \mathbb{C}^{N \times S}$ denote the equivalent channels from edge device k in cell l to AP l , from edge device k in cell l to the RIS, and from the RIS to AP l , respectively. We define the k -th edge device in \mathcal{K}_l as the (l, k) -th edge device. And then, the received signal at edge device k in \mathcal{K}_l from AP to the device and that from AP to RIS and to the device are given by

$$y_{l,k} = (\mathbf{h}_{l,k}^{rH} \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^{lH}) \mathbf{t}_l w_l + \sum_{j \neq l} (\mathbf{h}_{j,k}^{rH} \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{j,k}^{lH}) \mathbf{t}_j w_j + n_k, \quad (6)$$

where $\mathbf{t}_l \in \mathbb{C}^N$ denotes the transmit beamforming vector at AP l , and $n_k \sim \mathcal{CN}(0, \sigma_d^2)$ is the additive white Gaussian noise with zero mean and variance σ_d^2 at the (l, k) -th edge device. The transmit power constraint at AP l satisfies $E[|\mathbf{t}_l w_l|^2] = |\mathbf{t}_l|^2 \leq P_d$, where $P_d \geq 0$ denotes the maximum transmit power at AP l . Supposing perfect channel state information (CSI) is available, each edge device k in cell l can estimate the received global model by scaling a designed receive scalar $r_{l,k}$

which is set to $r_{l,k} = \left((\mathbf{h}_{l,k}^{rH} \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^{lH}) \mathbf{t}_l \right)^{-1}$. The received global model at edge device k is given by

$$w_{l,k} = r_{l,k} y_{l,k} = w_l + e_k^{\text{dl}}, \quad (7)$$

where $e_k^{\text{dl}} = \left(\sum_{j \neq l} (\mathbf{h}_{j,k}^{rH} \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{j,k}^{lH}) \mathbf{t}_j w_j + n_k \right) / \left((\mathbf{h}_{l,k}^{rH} \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^{lH}) \mathbf{t}_l \right)$ consists of the inter-cell interference and noise. Repeating d times, the global model is

$$\mathbf{w}_{l,k} = \mathbf{w}_l + \text{Re} \{ e_k^{\text{dl}} \}, \quad (8)$$

where $\mathbf{w}_{l,k}$, \mathbf{w}_l and e_k^{dl} are all vectors of dimension d . After receiving the global model $\mathbf{w}_{l,k}$, all edge devices start training based on the local data and then generate new local model parameters. The gradient information is the difference between the global model and the local model as in Eq. (4). After that, all edge devices upload their gradient information to AP l through the uplink communication.

2.4 Uplink AirComp Aggregation for RIS-Assisted FL System

In uplink communications, since the average sum in Eq. (5) for gradient aggregation is included in the category of nomographic functions, AirComp, as a promising technique, takes advantage of the waveform superposition properties of MACs in wireless networks to improve transmission efficiency. Fig. 2 shows the process of AirComp. For the sake of brevity, we also omit the time index in the following presentation. The transmitted signal and pre-processing function of the (l, k) -th edge device are denoted by $x_{l,k} \in \mathcal{C}$ and $\psi_{l,k}(\cdot): \mathcal{C} \rightarrow \mathcal{C}$, respectively. The target function processed at the l -th AP is given by

$$f = \phi \left(\sum_{k \in \mathcal{K}_l} \psi_{l,k}(x_{l,k}) \right), \quad (9)$$

where $\phi(\cdot)$ is the post-processing function at the AP. After pre-processing, the symbol transmitted at the (l, k) -th edge device $s_{l,k}$ is assumed to be independent and has the nature of zero mean and unit variance, i.e., $E[s_{l,k}] = 0$, $E[s_{l,k} s_{l,k}^H] = 1$. In this case, the aggregation at the l -th AP is expressed as

$$\mathbf{g}_l = \sum_{k \in \mathcal{K}_l} s_{l,k}. \quad (10)$$

Similar to the downlink communication, we let $\boldsymbol{\theta}_u = [\beta e^{j\theta_s^u}, \dots, \beta e^{j\theta_s^u}]$ represent the diagonal phase-shift matrix of the RIS in the uplink communication and $\boldsymbol{\Theta}_u = \text{diag}(\boldsymbol{\theta}_u)$ with $\theta_s^u \in [0, 2\pi]$. AP l mainly aggregates three types of signals, namely, the signal of cell l , the interference signal of other cells, and noise, where the first two items both contain the signal from the edge devices to AP l and the signal from the edge devices to

RIS and to AP l . Thus, the received signal at AP l is given by

$$\mathbf{y}_l = \sum_{k \in \mathcal{K}_l} (\mathbf{G}_l \Theta_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l) z_{l,k} s_{l,k} + \sum_{i \in \mathcal{K}_{j \neq l}} (\mathbf{G}_j \Theta_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) z_{j,i} s_{j,i} + \mathbf{n}_l, \quad (11)$$

where $z_{l,k} \in \mathcal{C}$ is the transmit scalar at the (l, k) -th edge device and it satisfies the maximum power constraint $|z_{l,k}|^2 \leq P$, and $\mathbf{n}_l \in \mathcal{C}^N \sim \mathcal{CN}(0, \sigma^2 I)$ denotes the additive white Gaussian noise with zero mean and variance σ^2 . Then, the estimated function at AP l after post-processing is marked as

$$\hat{g}_l = \frac{1}{\sqrt{\eta_l}} \mathbf{m}_l^H \mathbf{y}_l = \frac{1}{\sqrt{\eta_l}} \mathbf{m}_l^H \sum_{k \in \mathcal{K}_l} (\mathbf{G}_l \Theta_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l) z_{l,k} s_{l,k} + \frac{1}{\sqrt{\eta_l}} \mathbf{m}_l^H \sum_{i \in \mathcal{K}_{j \neq l}} (\mathbf{G}_j \Theta_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) z_{j,i} s_{j,i} + \frac{\mathbf{m}_l^H \mathbf{n}_l}{\sqrt{\eta_l}}, \quad (12)$$

where \mathbf{m}_l denotes the received beamforming vector at AP l and η_l denotes the denoising factor to suppress noise. Following Ref. [34], each transmit scalar can be designed as

$$z_{l,k} = \sqrt{\eta_l} (\mathbf{m}_l^H (\mathbf{G}_l \Theta_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l))^{-1}. \quad (13)$$

Therefore, the estimated function at AP l can further be expressed as

$$\hat{g}_l = g_l + e_l^{\text{ul}}, \quad (14)$$

where $e_l^{\text{ul}} = (\mathbf{m}_l^H / \sqrt{\eta_l}) \sum_{i \in \mathcal{K}_{j \neq l}} (\mathbf{G}_j \Theta_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) z_{j,i} s_{j,i} + (\mathbf{m}_l^H \mathbf{n}_l / \sqrt{\eta_l})$ denotes the total uplink error, which includes the inter-cell interference and noise. When AP l completes the aggregation process, a new round of global model updates is generated ac-

cording to Eq. (5), i.e., $\mathbf{w}_l^{t+1} = \mathbf{w}_l^t - \frac{\zeta^t}{K_l} \hat{g}_l$.

3 Convergence Analysis and Problem Formulation

In this section, we provide the convergence analysis of the proposed RIS-Assisted AirComp-based two-cell FL system. Based on the convergence results, we get an optimality gap bound that is influenced by both the downlink and uplink errors. In addition, we formulate the optimization problem to improve the performance of the proposed system.

3.1 Convergence Results

Assumption 1: M -Smoothness. All local loss functions (F_1, \dots, F_k) are M -Smoothness. For all \mathbf{x} and \mathbf{y} , we have

$$F_k(\mathbf{x}) \leq F_k(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^T \nabla F_k(\mathbf{x}) + \frac{M}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (15)$$

Assumption 2: μ -strongly convexity. All local loss functions F_1, \dots, F_k are μ -strongly convex. For all \mathbf{x} and \mathbf{y} , we have

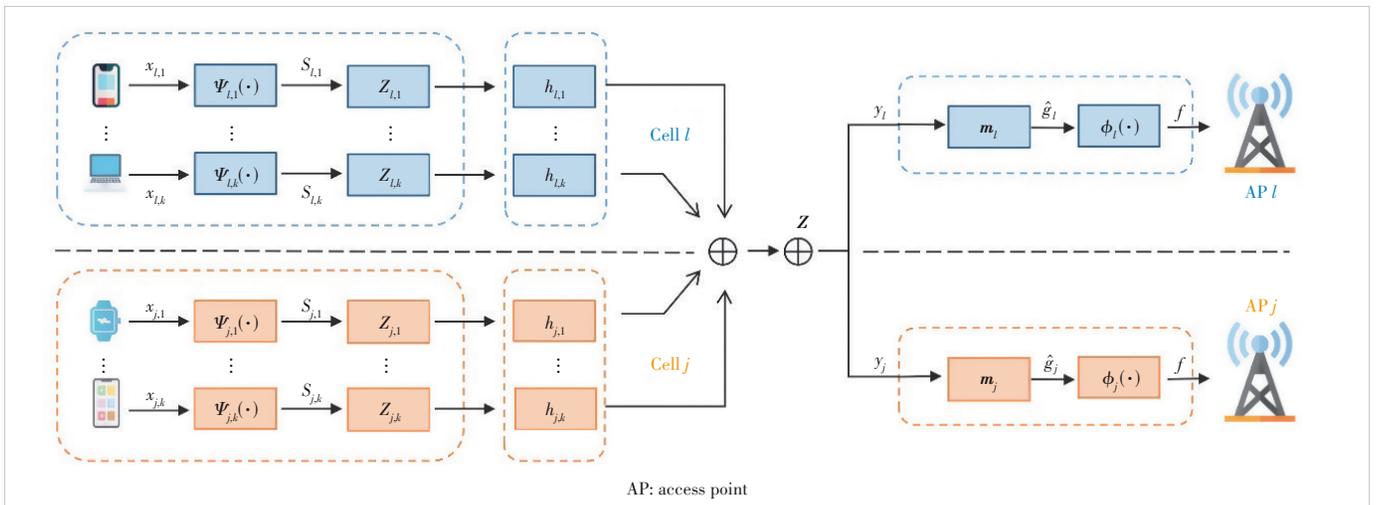
$$F_k(\mathbf{x}) \geq F_k(\mathbf{y}) + (\mathbf{x} - \mathbf{y})^T \nabla F_k(\mathbf{x}) + \frac{\mu}{2} \|\mathbf{x} - \mathbf{y}\|^2. \quad (16)$$

Theorem 1: Let Assumptions 1 and 2 be hold. In cell l , the learning rate satisfies $0 \leq \zeta_l \leq \zeta = 1/M$. After T communication rounds, the expected optimality gap in the RIS-Assisted FL system is upper bounded by

$$\mathbb{E}[F_l(\mathbf{w}_l^T) - F_l(\mathbf{w}_l^*)] \leq \rho^T \mathbb{E}[F_l(\mathbf{w}_l^0) - F_l(\mathbf{w}_l^*)] + \sum_{t=0}^{T-1} \rho^{T-t-1} \left(\frac{M}{2K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|e_{k,t}^{\text{dl}}\|^2] + \frac{1}{2MK^2} \mathbb{E}[\|e_{l,t}^{\text{ul}}\|^2] \right), \quad (17)$$

where $\rho = 1 - \mu/M$.

Proof: Please refer to Appendix for details.



▲ Figure 2. Process of over-the-air computation (AirComp) in the two-cell network

3.2 Problem Formulation

According to Theorem 1, the first term to the right of the inequality gradually tends to zero as the number of T increases. Thus, the upper bound is dominated by the last term, which includes the inter-cell interference and noise error in the downlink and uplink communications. We aim to minimize the upper bound in each time slot for transmitting the gradient information in all cells, given by

$$\sum_{l=1}^L \left(\frac{M}{2K} \sum_{k \in \mathcal{K}} \mathbb{E} \left[\left\| \mathbf{e}_{k,l}^{\text{dl}} \right\|^2 \right] + \frac{1}{2MK^2} \mathbb{E} \left[\left\| \mathbf{e}_{l,t}^{\text{ul}} \right\|^2 \right] \right), \forall t \in T, \forall l \in L, \quad (18)$$

We denote the optimization objective in Eq. (18) by the symbol \mathcal{E} . Specially, the denoising factor η_l in Eq. (14) is designed as

$$\eta_l = P \min_k \left\| \mathbf{m}_l^H (\mathbf{G}_l \boldsymbol{\Theta}_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l) \right\|^2. \quad (19)$$

Then, the corresponding optimization problem can be formulated as

$$\begin{aligned} & \text{minimize}_{\mathbf{m}_l, \boldsymbol{\Theta}_u, \boldsymbol{\Theta}_d, \mathbf{t}_l} \mathcal{E} \\ & \text{subject to } \left| \theta_s^{\text{ul}} \right| = 1, \forall s = 1, \dots, S, \\ & \quad \left| \theta_s^{\text{dl}} \right| = 1, \forall s = 1, \dots, S, \\ & \quad \left\| \mathbf{t}_l \right\|^2 \leq P_d, \end{aligned} \quad (20)$$

where θ_s^{ul} and θ_s^{dl} mean the phase-shift constraints, and \mathbf{t}_l is the transmit beamforming constraint.

$$\mathbb{E} \left[\left\| \mathbf{e}_{l,t}^{\text{dl}} \right\|^2 \right] = \frac{\sum_{j \neq l} \left\| (\mathbf{h}_{j,k}^H \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{j,k}^l) \mathbf{t}_j \right\|^2 + \sigma_d^2}{\left\| (\mathbf{h}_{l,k}^H \boldsymbol{\Theta}_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^l) \mathbf{t}_l \right\|^2}, \quad (21)$$

$$\mathbb{E} \left[\left\| \mathbf{e}_{l,t}^{\text{ul}} \right\|^2 \right] = \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} \frac{\eta_j \left\| \mathbf{m}_l^H (\mathbf{G}_j \boldsymbol{\Theta}_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) \right\|^2}{\eta_l \left\| \mathbf{m}_j^H (\mathbf{G}_j \boldsymbol{\Theta}_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) \right\|^2} + \frac{\left\| \mathbf{m}_l \right\|^2 \sigma^2}{\eta_l}. \quad (22)$$

For Problem (20), the optimization variables are the received beamforming vector \mathbf{m} , uplink phase-shift matrix $\boldsymbol{\Theta}_u$, transmit beamforming vector \mathbf{t} , and downlink phase-shift matrix $\boldsymbol{\Theta}_d$. The first two correspond to variables in the uplink process, and the last two are variables in the downlink process. We observe that the variables in these two processes are not coupled and their corresponding constraints are independent. Therefore, we can decompose the optimization objective into two sub-problems, i. e., downlink and uplink optimizations. Then, we can further solve Problem (20) by minimizing the following two sub-problems in Eqs. (23) and (24) simultaneously.

$$\begin{aligned} & \text{minimize}_{\mathbf{m}_l, \boldsymbol{\Theta}_u} \sum_{l=1}^L \mathbb{E} \left[\left\| \mathbf{e}_{l,t}^{\text{ul}} \right\|^2 \right], \\ & \text{subject to } \left| \theta_s^{\text{ul}} \right| = 1, \forall s = 1, \dots, S, \end{aligned} \quad (23)$$

$$\begin{aligned} & \text{minimize}_{\mathbf{t}_l, \boldsymbol{\Theta}_d} \sum_{l=1}^L \sum_{k \in \mathcal{K}} \mathbb{E} \left[\left\| \mathbf{e}_{k,l}^{\text{dl}} \right\|^2 \right], \\ & \text{subject to } \left| \theta_s^{\text{dl}} \right| = 1, \forall s = 1, \dots, S, \\ & \quad \left\| \mathbf{t}_l \right\|^2 \leq P_d. \end{aligned} \quad (24)$$

4 Optimization Framework

In this section, we specify the optimization framework for solving the uplink and downlink optimization problems, respectively. Each optimization problem also includes both beamforming optimization and phase-shift optimization.

4.1 Uplink Optimization

To simplify Eq. (19), we introduce an auxiliary variable vector $\boldsymbol{\gamma}_l = \min_k \left\| \mathbf{m}_l^H (\mathbf{G}_l \boldsymbol{\Theta}_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l) \right\|^2$ for cell l . By taking Eq. (19) to Problem (23) and introducing a new optimizing variable $\mathbf{v}_l = \mathbf{m}_l / \sqrt{\boldsymbol{\gamma}_l}$, the minimum problem in Eq. (23) can be adapted as

$$\begin{aligned} & \text{minimize}_{\mathbf{v}_l, \boldsymbol{\Theta}_u} \sum_{l=1}^L \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} \frac{\left\| \mathbf{v}_l^H (\mathbf{G}_j \boldsymbol{\Theta}_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) \right\|^2}{\left\| \mathbf{v}_j^H (\mathbf{G}_j \boldsymbol{\Theta}_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l) \right\|^2} + \sum_{l=1}^L q \left\| \mathbf{v}_l \right\|^2, \\ & \text{subject to } \left\| \mathbf{v}_l^H (\mathbf{G}_l \boldsymbol{\Theta}_u \mathbf{h}_{l,k}^r + \mathbf{h}_{l,k}^l) \right\|^2 \geq 1, \forall k \in \mathcal{K}_l, \forall l, \\ & \quad \left| \theta_s^{\text{ul}} \right| = 1, \forall s = 1, \dots, S. \end{aligned} \quad (25)$$

where $q = \sigma^2/P$ is a constant. We observe that the above problem turns out to be highly intractable due to the non-convexity of the objective function and nonconvex quadratic constraints for \mathbf{v} and $\boldsymbol{\Theta}$. First, we decompose the above optimization problem into $L + 1$ subproblems, i. e., L beamforming problems and a phase-shift problem. Then, we propose an alternative optimization algorithm to solve the uplink optimization problem.

1) Received beamforming optimization: We fix the diagonal phase-shift matrix $\boldsymbol{\Theta}_u$, and the l -th optimization sub-problem can be written as

$$\begin{aligned} & \text{minimize}_{\mathbf{v}_l} \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} \frac{\left\| \mathbf{v}_l^H \mathbf{h}_{j,i}^{l, \boldsymbol{\Theta}_u} \right\|^2}{\left\| \mathbf{v}_j^H \mathbf{h}_{j,i}^{j, \boldsymbol{\Theta}_u} \right\|^2} + q \left\| \mathbf{v}_l \right\|^2, \\ & \text{subject to } \left\| \mathbf{v}_l^H \mathbf{h}_{l,k}^{l, \boldsymbol{\Theta}_u} \right\|^2 \geq 1, \forall k \in \mathcal{K}_l, \forall l, \end{aligned} \quad (26)$$

where $\mathbf{h}_{j,i}^{l, \boldsymbol{\Theta}_u} = \mathbf{G}_j \boldsymbol{\Theta}_u \mathbf{h}_{j,i}^r + \mathbf{h}_{j,i}^l$, and $\mathbf{h}_{j,i}^{j, \boldsymbol{\Theta}_u}$ and $\mathbf{h}_{l,k}^{l, \boldsymbol{\Theta}_u}$ are also followed by this representation. Then we introduce an auxiliary

variable $b_{j,i}$ which satisfies $\frac{\|\mathbf{v}_l^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2}{\|\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2} \leq b_{j,i}$. Subsequently, Problem (26) can be equivalent to

$$\begin{aligned} & \underset{\mathbf{v}_l, b}{\text{minimize}} && \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} b_{j,i} + q \|\mathbf{v}_l\|^2, \\ & \text{subject to} && \|\mathbf{v}_l^H \mathbf{h}_{l,k}^{l,\Theta_u}\|^2 \geq 1, \forall k \in \mathcal{K}_l, \forall l, \\ & && \frac{\|\mathbf{v}_l^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2}{\|\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2} \leq b_{j,i}, \forall j \neq l. \end{aligned} \quad (27)$$

However, the constraints in Eq. (27) are nonconvex for the optimization variable \mathbf{v}_l . To address the nonconvexity of the constraints, we use the SCA method to transform the quadratic form into a linear constraint^[11]. We let $\mathbf{a}_{l,k} = [\text{Re}(\mathbf{v}_l^H \mathbf{h}_{l,k}^{l,\Theta_u}), \text{Im}(\mathbf{v}_l^H \mathbf{h}_{l,k}^{l,\Theta_u})]$ and $\mathbf{a}_{j,i} = [\text{Re}(\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}), \text{Im}(\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u})]$, and the corresponding linear constraints are

$$\begin{aligned} & \|\mathbf{a}_{l,k}^{(t)}\|^2 + 2(\mathbf{a}_{l,k}^{(t)})^T (\mathbf{a}_{l,k} - \mathbf{a}_{l,k}^{(t)}) \geq 1, \forall k, \forall l, \\ & \|\mathbf{a}_{j,i}^{(t)}\|^2 + 2(\mathbf{a}_{j,i}^{(t)})^T (\mathbf{a}_{j,i} - \mathbf{a}_{j,i}^{(t)}) \leq b_{j,i} \|\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2, \forall i, \forall j, \end{aligned} \quad (28)$$

where $\mathbf{a}_{l,k}^{(t)}$ and $\mathbf{a}_{j,i}^{(t)}$ are the t -th iteration solution. At the beginning of the iteration, the initial $\mathbf{a}_{l,k}^{(0)}$ and $\mathbf{a}_{j,i}^{(0)}$ can be randomly generated. By substituting Eq. (28) into Eq. (27), we have the following optimization problem:

$$\begin{aligned} & \underset{\mathbf{v}_l, b, \mathbf{a}}{\text{minimize}} && \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} b_{j,i} + q \|\mathbf{v}_l\|^2, \\ & \text{subject to} && \|\mathbf{a}_{l,k}^{(t)}\|^2 + 2(\mathbf{a}_{l,k}^{(t)})^T (\mathbf{a}_{l,k} - \mathbf{a}_{l,k}^{(t)}) \geq 1, \forall k, \forall l, \\ & && \|\mathbf{a}_{j,i}^{(t)}\|^2 + 2(\mathbf{a}_{j,i}^{(t)})^T (\mathbf{a}_{j,i} - \mathbf{a}_{j,i}^{(t)}) \leq B_{j,i}, \forall i, \forall j, \\ & && \mathbf{a}_{l,k} = [\text{Re}(\mathbf{v}_l^H \mathbf{h}_{l,k}^{l,\Theta_u}), \text{Im}(\mathbf{v}_l^H \mathbf{h}_{l,k}^{l,\Theta_u})], \\ & && \mathbf{a}_{j,i} = [\text{Re}(\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}), \text{Im}(\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u})], \end{aligned} \quad (29)$$

where $B_{j,i} = b_{j,i} \|\mathbf{v}_j^H \mathbf{h}_{j,i}^{l,\Theta_u}\|^2$. Then, we find that the objective function and constraints are convex for any optimization variable, which means we can adopt the CVX tools to obtain the optimal beamforming vector \mathbf{v}_l . When the beamforming vectors of all cells are obtained, we start optimizing the phase shift.

2) Uplink phase-shift optimization: With the given beamforming vector \mathbf{v} , we transform the channel as $\mathbf{G}_j \Theta_u \mathbf{h}_{j,i}^r = \mathbf{R}_{j,i}^r \boldsymbol{\theta}_u$, where $\boldsymbol{\theta}_u = \text{diag}(\Theta_u)$ and $\mathbf{R}_{j,i}^r \in \mathbb{C}^{N \times S}$ denotes the channel without phase-shift from node i to AP j . Then, the phase-shift optimization problem is rewritten as

$$\begin{aligned} & \underset{\boldsymbol{\theta}_u}{\text{minimize}} && \sum_{l=1}^L \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} \frac{\|\mathbf{v}_l^H (\mathbf{R}_{j,i}^r \boldsymbol{\theta}_u + \mathbf{h}_{j,i}^l)\|^2}{\|\mathbf{v}_j^H (\mathbf{R}_{j,i}^r \boldsymbol{\theta}_u + \mathbf{h}_{j,i}^j)\|^2}, \\ & \text{subject to} && \|\mathbf{v}_l^H (\mathbf{R}_{l,k}^r \boldsymbol{\theta}_u + \mathbf{h}_{l,k}^l)\|^2 \geq 1, \forall k \in \mathcal{K}_l, \forall l, \\ & && |\theta_s^{\text{ul}}| = 1, \forall s = 1, \dots, S. \end{aligned} \quad (30)$$

Unlike Problem (26), the optimization variable of the objective function in Problem (30) appears in both the numerator and denominator, which requires that we have to optimize the phase shift of all cells at the same time. For the equation constraints in Problem (30), we can reduce it to a convex constraint, i.e., $|\theta_s^{\text{ul}}| \leq 1$. In addition, we let $\mathbf{x}_{ji} = \mathbf{v}_l^H (\mathbf{R}_{j,i}^r \boldsymbol{\theta}_u + \mathbf{h}_{j,i}^l)$, $\mathbf{x}_{ji} = \mathbf{v}_j^H (\mathbf{R}_{j,i}^r \boldsymbol{\theta}_u + \mathbf{h}_{j,i}^j)$, $\mathbf{x}_{lk} = \mathbf{v}_l^H (\mathbf{R}_{l,k}^r \boldsymbol{\theta}_u + \mathbf{h}_{l,k}^l)$ and $\|\mathbf{x}_{ji}\|^2 / \|\mathbf{x}_{ji}\|^2 \leq r_{j,i}$. After applying the SCA method, the corresponding phase-shift problem is expressed as:

$$\begin{aligned} & \underset{\boldsymbol{\theta}_u, r, \mathbf{y}}{\text{minimize}} && \sum_{l=1}^L \sum_{\substack{i \in \mathcal{K}_j \\ j \neq l}} r_{j,i}, \\ & \text{subject to} && \|\mathbf{y}_{lk}^{(t)}\|^2 + 2(\mathbf{y}_{lk}^{(t)})^T (\mathbf{y}_{lk} - \mathbf{y}_{lk}^{(t)}) \geq 1, \forall k, \forall l, \\ & && \frac{\|\mathbf{y}_{ji}^{(t)}\|^2 + 2(\mathbf{y}_{ji}^{(t)})^T (\mathbf{y}_{ji} - \mathbf{y}_{ji}^{(t)})}{r_{j,i}} \leq \\ & && \|\mathbf{y}_{ji}^{(t)}\|^2 + 2(\mathbf{y}_{ji}^{(t)})^T (\mathbf{y}_{ji} - \mathbf{y}_{ji}^{(t)}) \leq r_{j,i}, \forall i, \forall j, \\ & && \mathbf{y}_{ji} = [\text{Re}(\mathbf{x}_{ji}), \text{Im}(\mathbf{x}_{ji})], \\ & && \mathbf{y}_{ji} = [\text{Re}(\mathbf{x}_{ji}), \text{Im}(\mathbf{x}_{ji})], \\ & && \mathbf{y}_{lk} = [\text{Re}(\mathbf{x}_{lk}), \text{Im}(\mathbf{x}_{lk})], \\ & && |\theta_s^{\text{ul}}| \leq 1, \forall s, \end{aligned} \quad (31)$$

where $\mathbf{y}_{lk}^{(t)}$, $\mathbf{y}_{ji}^{(t)}$ and $\mathbf{y}_{ji}^{(t)}$ are the t -th iteration solution. For Problem (31), the objective function and all constraints are convex, which indicates the optimal solution can be obtained from a convex program. Since we have scaled down the phase-shift equation constraints, when we get the optimal phase-shift solution from the convex program, we need to normalize it to satisfy the equation constraint.

The framework of optimization is summarized in Algorithm 2, where the process of solving Problems (29) and (31) is based on the SCA algorithm. For the equation constraint, we first relax it to obtain the optimal solution and then normalize the solution to satisfy the original condition.

Algorithm 2: Alternative beamforming and phase-shift algorithm

Input: The number of cells L , initial beamforming vector of each cell $\mathbf{v}_l, \in L$, initial random phase-shift matrix Θ_u , and

constant q .

Alternative beamforming optimization:

for the number of cell $l = 1, 2, \dots, L$ do

Fixing Θ_u and other cell $v_j, j \neq l$, introducing the auxiliary variable $b_{j,i}$;

$v_l \leftarrow$ solve Problem (29) by $(v_l, v_j, \Theta_u, b_{j,i})$;

end

Phase-shift algorithm:

Fixing the beamforming vector $v_l, \forall l$, introducing the auxiliary variable $z_{j,i}$;

Relaxing the equation constraint of Problem (30), i. e., $|\theta_s^{\text{ul}}| = 1$.

$\Theta_u \leftarrow$ solve Problem (31) by $(v_l, v_j, \Theta_u, z_{j,i})$;

$\Theta_u \leftarrow$ normalize Θ_u , i. e., $|\theta_s^{\text{ul}}| = |\theta_s^{\text{ul}}| / \text{abs}(\theta_s^{\text{ul}})$.

Output: $\{v_l, \forall l \in L, \Theta_u\}$.

4.2 Downlink Optimization

The downlink optimization problem is

$$\begin{aligned} & \underset{\mathbf{t}_l, \Theta_d}{\text{minimize}} && \sum_{l=1}^L \sum_{k \in \mathcal{K}} \frac{\sum_{j \neq l} \left\| \left(\theta_d^H \mathbf{T}_{j,k}^{r,H} + \mathbf{h}_{j,k}^{l,H} \right) \mathbf{t}_j \right\|^2 + \sigma_d^2}{\left\| \left(\mathbf{h}_{l,k}^{r,H} \Theta_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^{l,H} \right) \mathbf{t}_l \right\|^2}, \\ & \text{subject to} && \left\| \mathbf{t}_l \right\|^2 \leq P_d, \forall l, \\ & && \left| \theta_s^{\text{dl}} \right| = 1, \forall s = 1, \dots, S. \end{aligned} \quad (32)$$

We observe that Problem (32) is nonconvex for any optimization variable, and we cannot directly solve this optimization problem. For simplicity, we first let $\mathbf{h}_{j,k}^{l,\Theta_d} = \mathbf{h}_{j,k}^{r,H} \Theta_d \mathbf{G}_l^H + \mathbf{h}_{j,k}^{l,H}$ and $\mathbf{h}_{l,k}^{l,\Theta_d} = \mathbf{h}_{l,k}^{r,H} \Theta_d \mathbf{G}_l^H + \mathbf{h}_{l,k}^{l,H}$. Then we divide this optimization problem into two parts (transmit beamforming and downlink phase-shift optimizations).

1) Transmit beamforming optimization: For a given diagonal phase-shift matrix Θ_d , we mainly focus on the downlink received beamforming optimization. Then, we introduce an auxiliary variable $\Delta_{l,k}$ which satisfies $\left(\sum_{j \neq l} \left| \mathbf{h}_{j,k}^{l,\Theta_d} \mathbf{t}_j \right|^2 + \sigma_d^2 \right) / \left| \mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l \right|^2 \leq \Delta_{l,k}$. The optimization Problem (32) can now be converted to

$$\begin{aligned} & \underset{\{\mathbf{t}_l, \Delta\}}{\text{minimize}} && \sum_{l=1}^L \sum_{k \in \mathcal{K}} \Delta_{l,k}, \\ & \text{subject to} && \left\| \mathbf{t}_l \right\|^2 \leq P_d, \forall l, \\ & && \frac{\sum_{j \neq l} \left\| \mathbf{h}_{j,k}^{l,\Theta_d} \mathbf{t}_j \right\|^2 + \sigma_d^2}{\left\| \mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l \right\|^2} \leq \Delta_{l,k}, \forall k \in \mathcal{K}_l, \forall l \end{aligned} \quad (33)$$

For Constraint (33), we can adjust the inequality to $\left(d_{l,k} / \Delta_{l,k} \right) \leq \left\| \mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l \right\|^2$, where $d_{l,k} = \sum_{j \neq l} \left\| \mathbf{h}_{j,k}^{l,\Theta_d} \mathbf{t}_j \right\|^2 + \sigma_d^2$. In this case, for the nonconvex quadratic constraints concerning

the variable \mathbf{t}_l , we can exploit the SCA algorithm to linearly approximate the constraint as

$$\left\| \mathbf{c}_{l,k}^{(t)} \right\|^2 + 2(\mathbf{c}_{l,k}^{(t)})^T (\mathbf{c}_{l,k} - \mathbf{c}_{l,k}^{(t)}) \geq \frac{d_{l,k}}{\Delta_{l,k}}, \forall k, \forall l, \quad (34)$$

where $\mathbf{c}_{l,k} = [\text{Re}(\mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l), \text{Im}(\mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l)]$, $\forall k, \forall l$, and $\mathbf{c}_{l,k}^{(t)}$ is the optimized solution after the t -th iterative optimization. Then, the optimization problem at the l -th iteration is

$$\begin{aligned} & \underset{\mathbf{t}_l, \Delta}{\text{minimize}} && \sum_{l=1}^L \sum_{k \in \mathcal{K}} \Delta_{l,k}, \\ & \text{subject to} && \left\| \mathbf{t}_l \right\|^2 \leq P_d, \forall l, \\ & && \left\| \mathbf{c}_{l,k}^{(t)} \right\|^2 + 2(\mathbf{c}_{l,k}^{(t)})^T (\mathbf{c}_{l,k} - \mathbf{c}_{l,k}^{(t)}) \geq \frac{d_{l,k}}{\Delta_{l,k}}, \forall k, \forall l, \\ & && \mathbf{c}_{l,k} = \left[\text{Re}(\mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l), \text{Im}(\mathbf{h}_{l,k}^{l,\Theta_d} \mathbf{t}_l) \right], \forall k, \forall l \end{aligned} \quad (35)$$

Problem (35) is convex and we can easily solve it by utilizing convex optimization tools.

2) Downlink phase-shift optimization: We fix the transmit beamforming vector \mathbf{t} and denote the channel as $\mathbf{G}_l \Theta_d \mathbf{h}_{j,k}^r = \mathbf{T}_{j,k}^r \theta_d$, $\mathbf{G}_l \Theta_d \mathbf{h}_{l,k}^r = \mathbf{T}_{l,k}^r \theta_d$, where $\theta_d = \text{diag}(\Theta_d)$, $\mathbf{T}_{j,k}^r$ and $\mathbf{T}_{l,k}^r \in \mathbb{C}^{N \times S}$. The corresponding phase-shift optimization problem can be reformulated as

$$\begin{aligned} & \underset{\mathbf{t}_l, \Delta}{\text{minimize}} && \sum_{l=1}^L \sum_{k \in \mathcal{K}} \frac{\sum_{j \neq l} \left\| \left(\theta_d^H \mathbf{T}_{j,k}^{r,H} + \mathbf{h}_{j,k}^{l,H} \right) \mathbf{t}_j \right\|^2 + \sigma_d^2}{\left\| \left(\theta_d^H \mathbf{T}_{l,k}^{r,H} + \mathbf{h}_{l,k}^{l,H} \right) \mathbf{t}_l \right\|^2}, \\ & \text{subject to} && \left| \theta_s^{\text{dl}} \right| = 1, \forall s = 1, \dots, S. \end{aligned} \quad (36)$$

Problem (36) is in the same form as Problem (30), which means we can use the same strategy to solve the downlink phase-shift optimization.

5 Simulation Results

In this section, we provide some important simulation results to demonstrate the performance of the proposed RIS-assisted multi-cell FL network.

5.1 Experiment Setup

We consider a RIS-assisted two-cell wireless FL network in two-dimensional space where the coordinates of the APs are $(0, 0)$ and $(200, 0)$. The RIS is deployed at the edge of the two cells, i. e., $(100, 0)$. The edge devices of each cell are randomly scattered within a circle with a center of $(90, 0)$ and $(100, 0)$ and a radius of 10 m. We assume that the antennas of the APs and the reflecting elements of the RIS are both arranged in a uniform linear array. In the experiments, the path loss is modeled as $T(d/d_0)^{-\alpha}$ at a distance of $d_0 = 1$ m, where d denotes the link distance and α is the pass loss exponent.

We consider Rician fading for all channels and the channel coefficients are given as

$$\mathbf{h}_k = \sqrt{T(d_k/d_0)^{-\alpha}} \left(\sqrt{\frac{\beta}{1+\beta}} \mathbf{h}_k^{\text{LoS}} + \sqrt{\frac{\beta}{1+\beta}} \mathbf{h}_k^{\text{NLoS}} \right), \quad (37)$$

where $\mathbf{h}_k^{\text{LoS}}$ and $\mathbf{h}_k^{\text{NLoS}}$ represent the line-of-sight (LoS) and non-line-of-sight (NLoS) components. The Rician factor β is set to be 3. Particularly, we consider the same path loss exponent for all links, which is 2.2. Besides, we set $P_d = 30$ dBm, and $\sigma_d^2 = \sigma^2 = -10$ dBm, which means the constant $q = 1$.

In this paper, we adopt the sample-wise loss function and Modified National Institute of Standards and Technology (MNIST) datasets^[36] in the process of learning. We assume that each cell performs a different learning task (0 – 4 in Cell 1 and 5 – 9 in Cell 2) and that the learning rate is 0.1. The mini-batch datasets at different cells are 12 and 16, respectively. Next, we make the following specific schemes to compare the performance:

1) Without RIS: This scheme does not consider the RIS, which indicates the channel only contains the direct link between the APs and devices, i.e., $\Theta = 0$ (for both downlink and uplink communications).

2) Random phase-shift: Under this scheme, the phase-shift matrix is randomly generated in a RIS-assisted system, that is, we only need to optimize the beamforming vectors.

3) Optimal phase-shift: Under such a scheme, we optimize both the beamforming vectors and the phase-shift matrix of the RIS (Algorithm 2).

4) Error-free: The scheme is the benchmark of FL, which implies both the downlink model dissemination and uplink gradient aggregation are transmitted in an error-free manner.

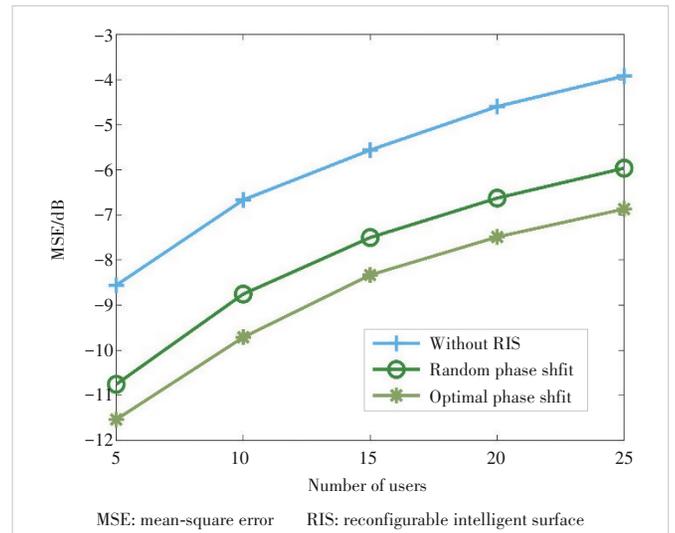
5.2 Performance of RIS-Assisted FL Two-Cell System

In this subsection, we first present the performance of the uplink aggregation based on AirComp and downlink dissemination error. Then we compare the performance of a two-cell FL system under different schemes.

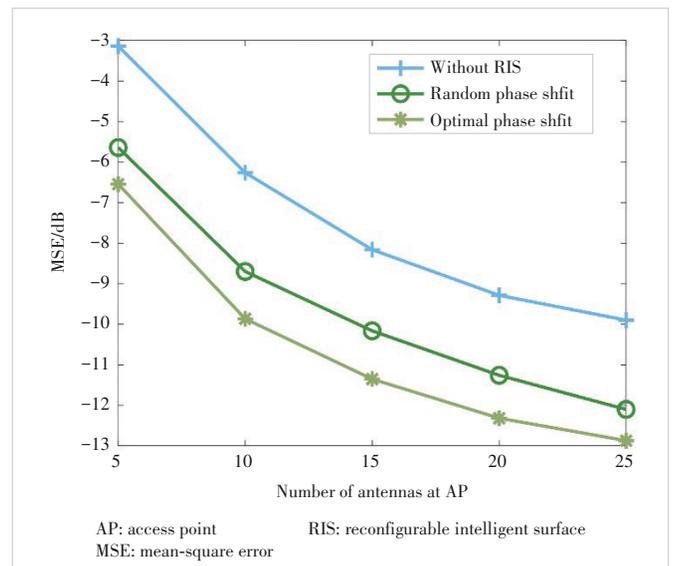
For the uplink aggregation, the mean-square error (MSE) is a very common performance metric in AirComp^[12, 14, 25, 34]. Therefore, we discuss the impact of the number of users, the number of antennas at each AP, and the number of reflecting elements at RIS on the average MSE across all cells. Fig. 3 displays the relationship between the MSE and the number of users, where the number of antennas at AP and the number of elements at RIS are set to be $N_1 = N_2 = 10$ and $S = 30$, respectively. It is obvious that the MSE increases with the number of users and deploying the RIS can significantly reduce the value of MSE compared to the absence of the RIS. This is because RIS can perform channel compensation for users at the edge of the corresponding cells with poor signals. On the one hand, with the increase of users, the inter-cell interference is more obvious, which

also enlarges the MSE. On the other hand, when a RIS is deployed at the edge of two cells, it can mitigate inter-cell interference. Besides, the RIS with optimal phase-shift is better than that RIS with random phase-shift on MSE, which indicates that the RIS with optimal phase-shift significantly enhances the signal strengths received at the APs. Fig. 4 compares the effects of the different numbers of antennas at AP on MSE, where the number of users per cell is fixed to 10 and the number of elements at RIS is also 30. We observe that the MSE decreases with the number of antennas, due to the diversity gain of antennas. RIS can improve the total MSE performance of the two-cell system. Correspondingly, the RIS with optimal phase shift can also achieve better MSE performance than the other two baseline schemes.

To compare the effect of the number of RIS elements on



▲ Figure 3. Relationship between MSE and the number of users



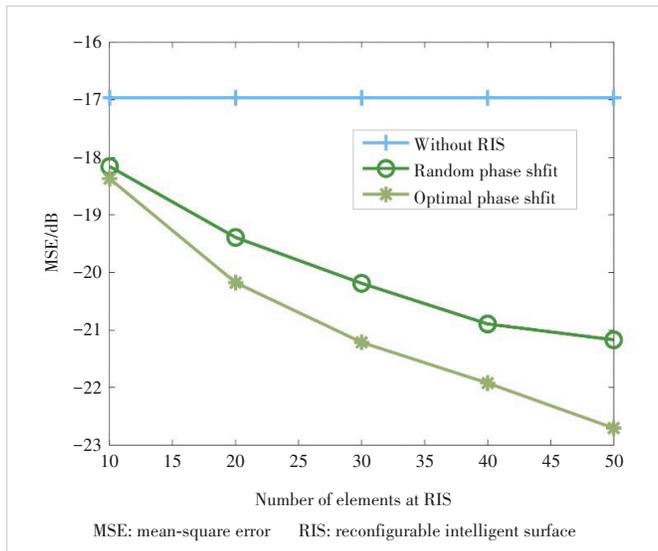
▲ Figure 4. Effects of the number of antennas on MSE

MSE, we first set the number of users and antennas at AP to 10, i.e., $N_1 = N_2 = K_1 = K_2 = 10$, and then we fix the location of users in each cell to avoid the influence of channel randomness. Fig. 5 shows that the number of elements at a RIS has a positive tendency correlated with the MSE, and as a result, the performance gradually gets better as the number of elements increases. In addition, the gap between random phase-shift and optimal phase-shift becomes larger and larger as the number of elements increases, which demonstrates the benefits of the optimal phase-shift scheme.

Since the downlink optimization and the uplink optimization have similar forms and are solved by the same algorithm, the impacts of the number of users and antennas at AP and the elements at the RIS on the downlink MSE have the same performance trend as those on the uplink MSE. We further compare the downlink errors in the case that $K_1 = K_2 = 10$, $N_1 = N_2 = 10$, and $S = 30$, i.e., $\sum_{k \in \mathcal{K}} \mathbb{E}[\|e_{k,d}^d\|^2]$. The results are shown in Table 1.

According to the results, the RIS with optimal phase shift still achieves the best performance, despite the small gaps in these errors. Moreover, we observe that the downlink error is much smaller than the uplink MSE, which indicates the downlink error has little effect on the convergence result of the overall system when the number of users is relatively small and $M = 10$ (the learning rate is $\zeta = 0.1$).

Next, we compare the performance of these schemes in the proposed two-cell FL system, where the number of users and that of antennas at AP in each cell are 5, and the number of el-



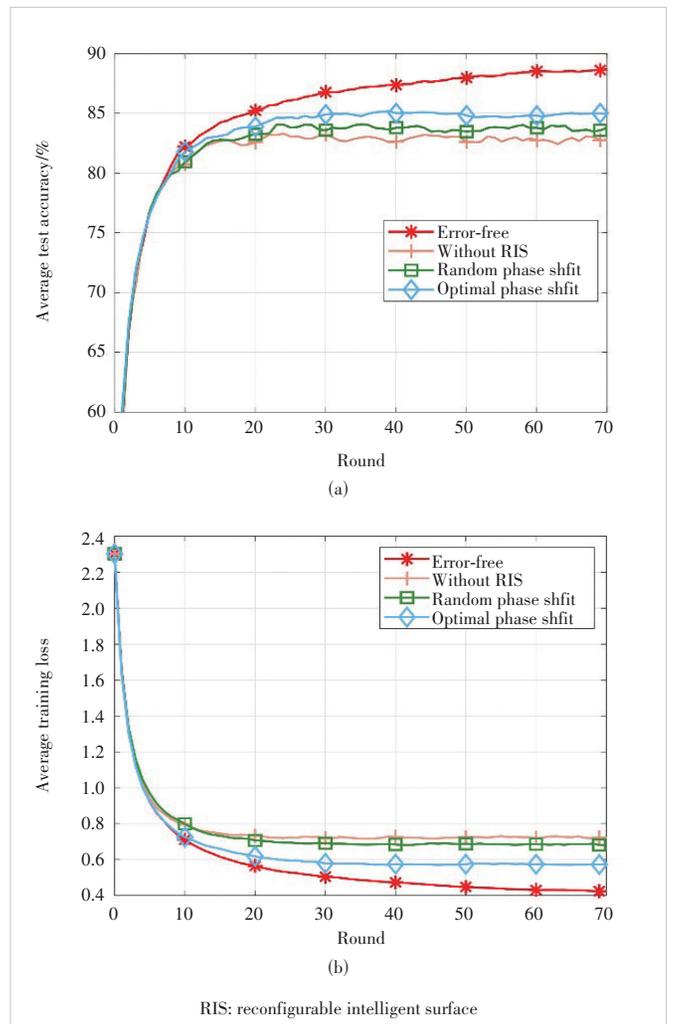
▲ Figure 5. Relationship between MSE and the number of elements at RIS

▼ Table 1. Comparison of downlink errors

Scheme	Error/dB
Without RIS	-52.77
Random PS	-53.16
Optimal PS	-53.91

PS: phase-shift RIS: reconfigurable intelligent surface

ements at RIS is set by 15. Each cell performs the same FL task with one local update in different mini-batch datasets. In order to compare the performance of the entire system, we average the train loss and test accuracy of the two cells and the results are shown in Fig. 6. Fig. 6 (a) shows, although the training loss of these schemes varies, all the schemes can achieve convergence and converge fast. Based on the proposed schemes, the RIS with optimal phase-shift scheme can demonstrate its advantages to enhance the performance of FL. From Fig. 6 (b), we notice that the RIS with optimal phase shift can achieve approximately 85% accuracy, the RIS with random phase-shift can get 83.5% accuracy, and the scheme without RIS only attains 82.7% accuracy, which proves that the RIS-assisted schemes can improve the performance of FL. To clearly show the effectiveness of our proposed system, we make additional time statistics for each scheme and each scheme runs for almost 800 s under $K = 5$, $M = 15$, $N = 5$, and $T = 300$, indicating that the proposed system can converge quickly. In



▲ Figure 6. Performance of different schemes in the proposed two-cell FL system: (a) training loss vs communication rounds; (b) test accuracy vs communication rounds

summary, RIS can compensate for the signal degradation of edge users and thereby decrease the error of communication. Moreover, we can adjust the phase-shift matrix of RIS to mitigate the inter-cell interference.

6 Conclusions and Future Work

In this paper, we develop a RIS-assisted AirComp-based two-cell FL wireless network, where each cell learns a different FL task and both the effects of downlink and uplink communications are considered. We first analyze the convergence of FL in the proposed system and show that the convergence is mainly influenced by the error of downlink and uplink transmissions. To enhance the performance of FL, we formulate the joint uplink and downlink optimization problem to minimize the optimality gap. To solve the problem, we divide the optimization problem into two separate subproblems. The beamforming vector and phase-shift matrix in each subproblem are optimized by alternative optimization based on SCA. In the end, simulation results show the performance and advantage of our proposed system and optimization algorithm.

In this work, we mainly focus on a scenario where a RIS assists two cells. In our future work, we will consider the scenario of a multi-RIS-assisted multi-cell wireless network, which makes the system model more complex. Since the placement of multi-RIS has a great impact on multi-cell performance, it is necessary to improve the average learning performance of all cells, as well as to avoid the poor performance of one cell. Most existing RISes only support a reflection or transmission mode. A new simultaneous transmitting and reflecting reconfigurable intelligent surface (STAR-RIS) can achieve full spatial coverage and have the advantage of adjusting more degrees of freedom. Therefore, promoting the deployment of SART-RIS is conducive to the implementation of more application scenarios.

Appendix

Proof of Theorem 1

For presentation clarity, we omit the cell index in the following analysis. According to Eqs. (5), (7) and (14), we have

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \frac{\zeta^t}{K} \left(\sum_{k \in \mathcal{K}_i} \nabla F_k(\mathbf{w}^t + \mathbf{e}_k^{\text{dl}}) + \mathbf{e}_i^{\text{ul}} \right). \quad (38)$$

Let $\nabla F(\hat{\mathbf{w}}^t) = \frac{1}{K} \sum_{k \in \mathcal{K}_i} \nabla F_k(\mathbf{w}^t + \mathbf{e}_{k,t}^{\text{dl}})$ and $\mathbf{e}_i^{\text{up}} = \frac{1}{K} \mathbf{e}_i^{\text{ul}}$, and then the global update can be rewritten by

$$\mathbf{w}^{t+1} = \mathbf{w}^t - \zeta_t (\nabla F(\hat{\mathbf{w}}^t) + \mathbf{e}_i^{\text{up}}). \quad (39)$$

According to Assumption 1, we obtain

$$F(\mathbf{w}^{t+1}) - F(\mathbf{w}^t) \leq \langle \nabla F(\mathbf{w}^t), \mathbf{w}^{t+1} - \mathbf{w}^t \rangle + \frac{M}{2} \|\mathbf{w}^{t+1} - \mathbf{w}^t\|^2 = \frac{M\zeta_t^2}{2} \|\nabla F(\hat{\mathbf{w}}^t) + \mathbf{e}_i^{\text{up}}\|^2 - \zeta_t \langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) + \mathbf{e}_i^{\text{up}} \rangle. \quad (40)$$

By taking the expectation on both sides of Eq. (40) and utilizing $\mathbb{E}[\mathbf{e}_i^{\text{up}}] = \mathbf{0}$, we have

$$\begin{aligned} \mathbb{E}[F(\mathbf{w}^{t+1}) - F(\mathbf{w}^t)] &\leq -\zeta_t \mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) + \mathbf{e}_i^{\text{up}} \rangle] + \\ &\frac{M\zeta_t^2}{2} \mathbb{E}[\|\nabla F(\hat{\mathbf{w}}^t) + \mathbf{e}_i^{\text{up}}\|^2] = \\ &-\zeta_t \mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) \rangle] + \frac{M\zeta_t^2}{2} \mathbb{E}[\|\nabla F(\hat{\mathbf{w}}^t)\|^2] + \\ &\frac{M\zeta_t^2}{2} \mathbb{E}[\|\mathbf{e}_i^{\text{up}}\|^2]. \end{aligned} \quad (41)$$

We let $T_1 = \mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) \rangle]$ and $T_2 = \mathbb{E}[\|\nabla F(\hat{\mathbf{w}}^t)\|^2]$. First, we make an upper bound of T_2 , and then we have

$$\begin{aligned} T_2 &= \mathbb{E}[\|\nabla F(\hat{\mathbf{w}}^t) \pm \nabla F(\mathbf{w}^t)\|^2] = \\ &\mathbb{E}[\|\nabla F(\hat{\mathbf{w}}^t) - \nabla F(\mathbf{w}^t)\|^2] + \mathbb{E}[\|\nabla F(\mathbf{w}^t)\|^2] + \\ &2\mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) - \nabla F(\mathbf{w}^t) \rangle] \leq \\ &\frac{M^2}{K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{e}_{k,t}^{\text{dl}}\|^2] - \mathbb{E}[\|\nabla F(\mathbf{w}^t)\|^2] + \\ &2\mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) \rangle], \end{aligned} \quad (42)$$

where $\mathbf{a} \pm \mathbf{b} = \mathbf{a} + \mathbf{b} - \mathbf{b}$ and the last inequality follows M -smoothness property $\|\nabla F(\mathbf{x}) - \nabla F(\mathbf{y})\| \leq M\|\mathbf{x} - \mathbf{y}\|$. Therefore,

$$\begin{aligned} \mathbb{E}[F(\mathbf{w}^{t+1}) - F(\mathbf{w}^t)] &\leq -\zeta_t (1 - M\zeta_t) \mathbb{E}[\langle \nabla F(\mathbf{w}^t), \nabla F(\hat{\mathbf{w}}^t) \rangle] + \\ &\frac{M^3\zeta_t^2}{2K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{e}_{k,t}^{\text{dl}}\|^2] + \frac{M\zeta_t^2}{2} \mathbb{E}[\|\mathbf{e}_i^{\text{up}}\|^2] - \frac{M\zeta_t^2}{2} \mathbb{E}[\|\nabla F(\mathbf{w}^t)\|^2]. \end{aligned} \quad (43)$$

By setting $0 \leq \zeta_t \equiv \zeta = \frac{1}{M}$, we obtain

$$\begin{aligned} \mathbb{E}[F(\mathbf{w}^{t+1}) - F(\mathbf{w}^t)] &\leq -\frac{1}{2M} \mathbb{E}[\|\nabla F(\mathbf{w}^t)\|^2] + \\ &\frac{M}{2K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{e}_{k,t}^{\text{dl}}\|^2] + \frac{1}{2M} \mathbb{E}[\|\mathbf{e}_i^{\text{up}}\|^2]. \end{aligned} \quad (44)$$

Based on Assumption 2, we have $\|\nabla F(\mathbf{w}^t)\|^2 \geq 2\mu(F(\mathbf{w}^t) - F(\mathbf{w}^*))$.

Thus, Eq. (44) can be represented as

$$\begin{aligned} \mathbb{E}[F(\mathbf{w}^{t+1}) - F(\mathbf{w}^t)] &\leq -\frac{\mu}{M} \mathbb{E}[F(\mathbf{w}^t) - \\ &F(\mathbf{w}^*)] + \frac{M}{2K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|\mathbf{e}_{k,t}^{\text{dl}}\|^2] + \frac{1}{2M} \mathbb{E}[\|\mathbf{e}_i^{\text{up}}\|^2]. \end{aligned} \quad (45)$$

Rearranging Eq. (45) and applying recursion, we have

$$\mathbb{E}[F(\mathbf{w}^T) - F(\mathbf{w}^*)] \leq \rho^T \mathbb{E}[F(\mathbf{w}^0) - F(\mathbf{w}^*)] + \sum_{t=0}^{T-1} \rho^{T-t-1} \left(\frac{M}{2K} \sum_{k \in \mathcal{K}} \mathbb{E}[\|e_{k,d}^{\text{dl}}\|^2] + \frac{1}{2MK^2} \mathbb{E}[\|e_t^{\text{ul}}\|^2] \right), \quad (46)$$

where $\rho = 1 - \mu/M$. Therefore we get Theorem 1.

References

- [1] LETAIEF K B, SHI Y M, LU J M, et al. Edge artificial intelligence for 6G: vision, enabling technologies, and applications [J]. *IEEE journal on selected areas in communications*, 2021, 40(1): 5 - 36. DOI: 10.1109/JSAC.2021.3126076
- [2] LETAIEF K B, CHEN W, SHI Y M, et al. The roadmap to 6G: AI empowered wireless networks [J]. *IEEE communications magazine*, 2019, 57(8): 84 - 90. DOI: 10.1109/mcom.2019.1900271
- [3] YANG Q, LIU Y, CHEN T J, et al. Federated machine learning: concept and applications [J]. *ACM transactions on intelligent systems and technology*, 2019, 10(2): 1 - 19. DOI: 10.1145/3298981
- [4] WEN D Z, JEON K J, HUANG K B. Federated dropout—a simple approach for enabling federated learning on resource constrained devices [J]. *IEEE wireless communications letters*, 2022, 11(5): 923 - 927. DOI: 10.1109/LWC.2022.3149783
- [5] MCMAHAN H B, MOORE E, RAMAGE D, et al. Communication-efficient learning of deep networks from decentralized data [C]//20th International Conference on Artificial Intelligence and Statistics (AISTATS). *JMLR*, 2017: 1273 - 1282. DOI: 10.48550/arXiv.1602.05629
- [6] XU W, YANG Z H, NG D W K, et al. Edge learning for 5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing [J]. *IEEE journal of selected topics in signal processing*, 2023: 1 - 31. DOI: 10.1109/jstsp.2023.3239189
- [7] CHEN M Z, YANG Z H, SAAD W, et al. A joint learning and communications framework for federated learning over wireless networks [J]. *IEEE transactions on wireless communications*, 2021, 20(1): 269 - 283. DOI: 10.1109/TWC.2020.3024629
- [8] YANG H H, LIU Z Z, QUEK T Q S, et al. Scheduling policies for federated learning in wireless networks [J]. *IEEE transactions on communications*, 2020, 68(1): 317 - 333. DOI: 10.1109/tcomm.2019.2944169
- [9] YANG P, JIANG Y N, WANG T, et al. Over-the-air federated learning via second-order optimization [J]. *IEEE transactions on wireless communications*, 2022, 21(12): 10560 - 10575. DOI: 10.1109/TWC.2022.3185156
- [10] NAZER B, GASTPAR M. Computation over multiple-access channels [J]. *IEEE transactions on information theory*, 2007, 53(10): 3498 - 3516. DOI: 10.1109/tit.2007.904785
- [11] CHEN L, QIN X W, WEI G. A uniform-forcing transceiver design for over-the-air function computation [J]. *IEEE wireless communications letters*, 2018, 7(6): 942 - 945. DOI: 10.1109/LWC.2018.2840157
- [12] YANG Y H, ZHOU Y, WU Y L, et al. Differentially private federated learning via reconfigurable intelligent surface [J]. *IEEE Internet of Things journal*, 2022, 9(20): 19728 - 19743. DOI: 10.1109/JIOT.2022.3168066
- [13] WANG Z B, ZHAO Y P, ZHOU Y, et al. Over-the-air computation: foundations, technologies, and applications [EB/OL]. [2022-10-19]. <https://arxiv.org/abs/2210.10524>
- [14] YANG K, JIANG T, SHI Y M, et al. Federated learning via over-the-air computation [J]. *IEEE transactions on wireless communications*, 2020, 19(3): 2022 - 2035. DOI: 10.1109/TWC.2019.2961673
- [15] FANG W Z, YU Z Y, JIANG Y N, et al. Communication-efficient stochastic zero-order optimization for federated learning [J]. *IEEE transactions on signal processing*, 2022, 70: 5058 - 5073. DOI: 10.1109/TSP.2022.3214122
- [16] FU M, SHI Y M, ZHOU Y. Federated learning via unmanned aerial vehicle [EB/OL]. [2022-10-20]. <https://arxiv.org/abs/2210.10970>
- [17] LIM W Y B, GARG S, XIONG Z H, et al. UAV-assisted communication efficient federated learning in the era of the artificial intelligence of things [J]. *IEEE network*, 2021, 35(5): 188 - 195. DOI: 10.1109/MNET.002.2000334
- [18] NI W L, LIU Y W, YANG Z H, et al. Federated learning in multi-RIS-aided systems [J]. *IEEE Internet of Things journal*, 2022, 9(12): 9608 - 9624. DOI: 10.1109/JIOT.2021.3130444
- [19] WANG Z B, QIU J H, ZHOU Y, et al. Federated learning via intelligent reflecting surface [J]. *IEEE transactions on wireless communications*, 2022, 21(2): 808 - 822. DOI: 10.1109/twc.2021.3099505
- [20] YANG K, SHI Y M, ZHOU Y, et al. Federated machine learning for intelligent IoT via reconfigurable intelligent surface [J]. *IEEE network*, 2020, 34(5): 16 - 22. DOI: 10.1109/MNET.011.2000045
- [21] LIU H, YUAN X J, ZHANG Y J A. Joint communication-learning design for RIS-assisted federated learning [C]//IEEE International Conference on Communications Workshops (ICC Workshops). *IEEE*, 2021: 1 - 6. DOI: 10.1109/ICCWorkshops50388.2021.9473672
- [22] YIN T, LI L X, MA D H, et al. FLIGHT: federated learning with IRS for grouped heterogeneous training [J]. *Journal of communications and information networks*, 2022, 7(2): 135 - 144. DOI: 10.23919/jcin.2022.9815197
- [23] HU L, WANG Z B, ZHU H B, et al. RIS-assisted over-the-air federated learning in millimeter wave MIMO networks [J]. *Journal of communications and information networks*, 2022, 7(2): 145 - 156. DOI: 10.23919/jcin.2022.9815198
- [24] WU Q Q, ZHANG R. Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming [J]. *IEEE transactions on wireless communications*, 2019, 18(11): 5394 - 5409. DOI: 10.1109/TWC.2019.2936025
- [25] FANG W Z, JIANG Y N, SHI Y M, et al. Over-the-air computation via reconfigurable intelligent surface [J]. *IEEE transactions on communications*, 2021, 69(12): 8612 - 8626. DOI: 10.1109/tcomm.2021.3114791
- [26] HUANG C W, ZAPPONE A, ALEXANDROPOULOS G C, et al. Reconfigurable intelligent surfaces for energy efficiency in wireless communication [J]. *IEEE transactions on wireless communications*, 2019, 18(8): 4157 - 4170. DOI: 10.1109/TWC.2019.2922609
- [27] WEINBERGER K, AHMAD A A, SEZGIN A, et al. Synergistic benefits in IRS- and RS-enabled C-RAN with energy-efficient clustering [J]. *IEEE transactions on wireless communications*, 2022, 21(10): 8459 - 8475. DOI: 10.1109/TWC.2022.3166393
- [28] ZHAI X F, HAN G J, CAI Y L, et al. Joint beamforming aided over-the-air computation systems relying on both BS-side and user-side reconfigurable intelligent surfaces [J]. *IEEE transactions on wireless communications*, 2022, 21(12): 10766 - 10779. DOI: 10.1109/TWC.2022.3187156
- [29] WANG Z B, ZHOU Y, SHI Y M, et al. Interference management for over-the-air federated learning in multi-cell wireless networks [J]. *IEEE journal on selected areas in communications*, 2022, 40(8): 2361 - 2377. DOI: 10.1109/JSAC.2022.3180799
- [30] PAN C H, REN H, WANG K Z, et al. Multicell MIMO communications relying on intelligent reflecting surfaces [J]. *IEEE transactions on wireless communications*, 2020, 19(8): 5218 - 5233. DOI: 10.1109/twc.2020.2990766
- [31] LUO C H, LI X, JIN S, et al. Reconfigurable intelligent surface-assisted multi-cell MISO communication systems exploiting statistical CSI [J]. *IEEE wireless communications letters*, 2021, 10(10): 2313 - 2317. DOI: 10.1109/LWC.2021.3100427
- [32] XIE H L, XU J, LIU Y F. Max-min fairness in IRS-aided multi-cell MISO systems via joint transmit and reflective beamforming [C]//IEEE International Conference on Communications (ICC). *IEEE*, 2020: 1 - 6. DOI: 10.1109/ICC40277.2020.9148858
- [33] NI W L, LIU X, LIU Y W, et al. Resource allocation for multi-cell IRS-aided NOMA networks [J]. *IEEE transactions on wireless communications*, 2021, 20(7): 4253 - 4268. DOI: 10.1109/TWC.2021.3057232
- [34] LI J, FU M, ZHOU Y, et al. Double-RIS assisted over-the-air computation [C]//IEEE Globecom Workshops (GC Wkshps). *IEEE*, 2022: 1 - 6. DOI: 10.1109/GCWkshps52748.2021.9682077
- [35] ABARI O, RAHUL H, KATABI D, et al. AirShare: distributed coherent transmission made seamless [C]//IEEE Conference on Computer Communications (INFOCOM). *IEEE*, 2015: 1742 - 1750. DOI: 10.1109/INFOCOM.2015.7218555

- [36] LECUN Y, BOTTOU L, BENGIO Y, et al. Gradient-based learning applied to document recognition [J]. *Proceedings of the IEEE*, 1998, 86(11): 2278 - 2324. DOI: 10.1109/5.726791

Biographies

WANG Yiji received his BS degree from Zhejiang University City College, China in 2020. He is currently pursuing his master's degree with the School of Information Science and Technology, ShanghaiTech University, China. His research interests include federated learning and wireless communications.

WEN Dingzhu (wendzh@shanghaitech.edu.cn) received his bachelor's and master's degrees from Zhejiang University, China in 2014 and 2017, respectively, and PhD degree from The University of Hong Kong, China in 2021. Subsequently, he joined ShanghaiTech University, China. He is currently an assistant professor at the School of Information Science and Technology there. His research interests include edge intelligence, integrated sensing, computation and communication, over-the-air computation, in-band full-duplex communications, etc.

MAO Yijie is an assistant professor at the School of Information Science and Technology, ShanghaiTech University, China. She received her BE degree from Beijing University of Posts and Telecommunications, China and BE

(Hons.) degree from the Queen Mary University of London in 2014. She received her PhD degree from the Electrical and Electronic Engineering (EEE) Department, The University of Hong Kong, China in 2018. She was a postdoctoral research fellow at The University of Hong Kong from 2018 to 2019 and a postdoctoral research associate with the Department of the EEE at the Imperial College London, UK from 2019 to 2021. She is a senior member of China Institute of Communications. She is currently serving as an editor of *IEEE Communications Letters* and a guest editor of two special issues of *IEEE Journal on Selected Areas in Communications* and *IEEE Open Journal of the Communications Society*.

SHI Yuanming received his BS degree in electronic engineering from Tsinghua University, China in 2011. He received his PhD degree in electronic and computer engineering from The Hong Kong University of Science and Technology (HKUST), China in 2015. Since September 2015, he has been with the School of Information Science and Technology, ShanghaiTech University, China, where he is currently a tenured associate professor. He visited University of California, Berkeley, USA from October 2016 to February 2017. His research areas include optimization, machine learning, wireless communications, and their applications to 6G, IoT and edge AI. He was a recipient of the 2016 IEEE Marconi Prize Paper Award in Wireless Communications, the 2016 Young Author Best Paper Award by the IEEE Signal Processing Society, and the 2021 IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award. He is also an editor of *IEEE Transactions on Wireless Communications*, *IEEE Journal on Selected Areas in Communications*, and *Journal of Communications and Information Networks*.