Latency-Aware Downlink Mutual Information Maximization for RIS-Assisted Federated Learning

Yujun Cai, *Graduate Student Member, IEEE*, Shufeng Li, *Member, IEEE*, Jianbo Liu, Shaolin Liao, *Senior Member, IEEE*, and Xinruo Zhang, *Member, IEEE*

***Abstract*—Federated learning (FL) has emerged as an important framework for distributed edge intelligence over wireless networks, where the wall-clock training efficiency is strongly affected by the communication latency for global model dissemination and aggregation. In reconfigurable intelligent surface (RIS)-assisted wireless systems, improving the downlink transmission quality can support more timely model delivery and, consequently, faster FL progress under a fixed training-time budget. In this paper, we study a multi-user RIS-assisted FL downlink and develop an FL-oriented communication model that relates the global-model dissemination latency to the achievable mutual information (MI). To account for the bottleneck nature of FL downlink transmission, we clarify the relationship between sum-MI maximization and worst-user latency through a rate-dispersion factor, under which sum MI serves as a tractable communication-side surrogate for improving wall-clock FL efficiency. Based on this model, we formulate a joint active and passive beamforming problem under the base station transmit-power constraint and the RIS unit-modulus constraint. To handle the resulting non-convex coupling, we develop an efficient alternating optimization (AO) framework that combines weighted minimum mean-square error based active beamforming and block coordinate descent based passive beamforming, with monotonic objective improvement under the available channel state information (CSI). The impact of imperfect CSI is also discussed to reflect practical RIS-assisted deployment. Simulation results show that the proposed design achieves noticeable MI gains over representative communication-learning benchmarks, which are empirically associated with reduced downlink dissemination latency, improved communication–computation efficiency, and faster FL progress on standard image classification tasks.**

***Index Terms*—Reconfigurable intelligent surface, federated learning, beamforming, mutual information**

I. Introduction

The contemporary digital landscape is characterized by the widespread proliferation of edge computing devices, which continuously generate vast quantities of sensitive user data. In bandwidth-limited wireless federated learning (FL) systems, communication latency is a major bottleneck to wall-clock training efficiency. This motivates communication-aware physical-layer design that can improve the timeliness of model dissemination and support faster learning progress under a fixed training time budget.

FL is an emerging paradigm that enables training machine learning models directly on users' devices while keeping data private. In FL, multiple devices collaboratively train a shared model using their local datasets. Each device updates the model locally and sends only the model parameters to a central server without raw data. The server aggregates these updates to produce an improved global model, which is then redistributed to all devices for further refinement [1].

Recently, researchers have made substantial progress on communication-aware FL and reconfigurable intelligent surface (RIS)-assisted wireless transmission. Existing studies have investigated robust transceiver design, over-the-air computation, device scheduling, latency-aware resource allocation, and RIS-assisted beamforming for wireless FL. However, most of these works either focus on uplink aggregation reliability, communication–computation co-design, or conventional communication metrics such as sum rate and capacity. Comparatively less attention has been paid to FL-oriented downlink dissemination design in RIS-assisted systems, especially when the communication objective is explicitly linked to wall-clock training efficiency. In practice, downlink can be equally critical, especially in modern settings for large-scale FL systems and when large models are to be disseminated.

In this paper, we study a multi-user RIS-assisted FL downlink, where the base station (BS) disseminates the global model to multiple devices through parallel downlink transmission. To ensure model consistency, all participant devices start local training from the same global model; hence, the weakest user becomes the bottleneck. Different from conventional RIS sum-rate maximization formulations, our goal is not merely to improve throughput in a generic communication sense, but to develop an FL-oriented communication model in which physical-layer performance is connected to global-model dissemination efficiency under a fixed training time budget. Based on this perspective, we formulate a joint active and passive beamforming problem that maximizes the system sum mutual information (MI) as a tractable communication-side surrogate, and show that the resulting design is empirically associated with lower dissemination latency and faster FL progress.

The main contributions of this paper are summarized as follows:

1) We develop an FL-oriented communication model for a

This work was supported by the National Natural Science Foundation of China No. 62371428, the Royal Society International Exchanges 2023 Cost Share No. IEC/NSFC/233318, and the Fundamental Research Funds for the Central Universities No. CUC26TD06. *(Corresponding author: Shufeng Li, Shaolin Liao)*

Yujun Cai, Shufeng Li, and Jianbo Liu are with the School of Information and Communication Engineering, Communication University of China, Beijing 100024, China (e-mail: tangentchase@cuc.edu.cn; lishufeng@cuc.edu.cn; ljb@cuc.edu.cn).

Shaolin Liao is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou 510275, China, the Department of Electrical and Computer Engineering, Illinois Institute of Technology, Chicago, IL, USA 60616, and also with Elmore School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN, USA 47907 (e-mail: liaoshlin@mail.sysu.edu.cn).

Xinruo Zhang is with the School of Computer Science and Electronic Engineering, University of Essex, Colchester CO4 3SQ, United Kingdom (e-mail: xinruo.zhang@essex.ac.uk).

multi-user RIS-assisted downlink and relate the achievable physical-layer MI to the global-model dissemination latency. Under a fixed training time budget, this connection motivates sum-MI maximization as a tractable communication-side design objective for improving wall-clock FL efficiency.

2) We formulate a joint active and passive beamforming design problem for the considered RIS-assisted FL downlink, where the BS beamformers and RIS phase shifts are optimized under the BS transmit power constraint and the RIS unit-modulus constraint. The resulting problem captures the coupling between wireless downlink design and model dissemination efficiency in FL systems.
3) To solve the non-convex problem efficiently, we develop an alternating optimization framework that combines a weighted minimum mean-square error (WMMSE)-based active beamforming update with a low-complexity block coordinate descent (BCD)-based passive beamforming update. The proposed method ensures monotonic objective improvement and has practical computational complexity. Simulation results demonstrate consistent MI gains over representative communication-learning benchmark pipelines, which are empirically associated with reduced downlink dissemination latency and faster FL progress.

The remainder of this paper is organized as follows. Section II reviews the related work and summarizes the distinctions of the proposed study. Section III introduces the system model and formulates the optimization problem. The proposed algorithm is presented in Section IV. Section V offers numerical results to demonstrate the performance of the proposed algorithm. Finally, Section VI concludes the paper.

II. Related Work

A. Wireless FL and Latency-Aware Design

Researchers have studied several aspects of wireless FL. Qi et al. [2] investigated the impact of wireless channel uncertainties on FL in edge-intelligent networks and proposed a robust worst-case optimized FL framework with joint device selection and transceiver design to ensure reliable and accurate model training. Sun et al. [3] proposed an energy-aware dynamic device scheduling algorithm for over-the-air federated edge learning that jointly accounts for communication and computation energy, leveraged gradient-norm estimation under uncertainty, and significantly improved training accuracy under energy constraints. Kim et al. [4] formulated and solved an over-the-air computation (AirComp) beamforming and device selection problem for FL by leveraging an AirComp–multicasting duality, enabling a low-complexity algorithm that maximizes participating devices under aggregation error constraints and achieves near-ideal learning performance.

Existing studies have also investigated latency-aware wireless FL from different perspectives. Park et al. [5] considered FL over a fog radio access network and proposed a rate-splitting transmission strategy that jointly optimizes communication and training parameters to minimize completion time under capacity constraints. Xu [6] proposed a time division multiple access (TDMA)-based wireless FL scheme that jointly optimizes user scheduling, time allocation, and resource management to minimize training latency under heterogeneous user constraints. Deng et al. [7] developed an efficient communication–computation FL framework for resource-constrained industrial networks by integrating deep neural networks partitioning with device-specific participation rates and jointly optimizing scheduling and resource allocation to minimize training delay. These studies collectively highlight the importance of latency-aware communication design for practical FL deployments.

B. MI-Driven FL and RIS-Assisted Communications

The interplay between FL and MI has also attracted research interest. Alsulaimawi [8] addressed data security issues by introducing an MI-driven parameter aggregation algorithm, which bolstered model accuracy and resilience against adversarial attacks. Zhang et al. [9] introduced FedMIC for wireless traffic prediction, where an MI-based spectral clustering algorithm and a hierarchical aggregation framework with attention were adopted to enhance prediction precision and generalization performance. Chen et al. [10] proposed an MI-based FL aggregation algorithm that assigned trust weights to devices by measuring gradient consistency, effectively mitigating malicious node attacks and improving robustness and accuracy in edge computing networks.

Beyond FL, RIS has emerged as a promising wireless communication paradigm for improving signal quality and network performance through the adaptive manipulation of electromagnetic wave propagation within wireless channels [11]. Extensive studies have explored RIS-assisted transmission in physical-layer security, non-orthogonal multiple access (NOMA), and multiple-input multiple-output (MIMO) communications [12], [13], [14]. Ma et al. [15] showed that jointly considering physical-layer beamforming and semantic-layer network training in an RIS-assisted system can yield substantial performance improvements. Zhang et al. [16] studied integrated sensing and communication, where unified waveform and phase-shift design enhances both communication and sensing MI. Tuan et al. [17] further developed scalable algorithms to evaluate system sum rate and optimize the minimum user rate in RIS-enhanced broadcast channels. Along this line, Guo et al. [18] studied weighted sum-rate maximization for RIS-assisted wireless networks, which serves as a representative communication-centric RIS beamforming work.

Advanced RIS architectures have also been explored. Han et al. [19] analyzed a cooperative dual-RIS design and showed that jointly optimizing the transmitter and two RISs can effectively boost capacity in line-of-sight scenarios. Lee et al. [20] introduced RIS phase-shifter-based virtual multipath channels to enhance time-reversal multiple access, enabling better support for NOMA users with improved capacity, outage, spectral efficiency, and lower complexity than conventional RIS-NOMA. Zhang et al. [21] quantified how non-ideal channel conditions influence the capacities of primary and secondary transmissions in RIS-assisted symbiotic radio systems. More closely related to FL, Wang et al. [22] investigated

TABLE I
COMPARISON WITH REPRESENTATIVE RELATED WORK

Reference	Main focus	FL-aware	RIS-assisted	Joint design	Transmission stage
[4]	Beamforming vector design and device selection in over-the-air FL	✓		✓	Uplink aggregation
[5]	Rate-splitting design for fog radio access network (RAN) FL completion-time minimization	✓			Downlink/uplink system-level design
[6]	Time division multiple access (TDMA)-based latency minimization for wireless FL	✓			Communication scheduling and resource allocation
[7]	Communication–computation co-design for low-latency FL	✓			End-to-end FL system design
[18]	Weighted sum-rate maximization in RIS-assisted wireless networks		✓	✓	Generic downlink communication
[22]	RIS-assisted over-the-air FL with device scheduling and beamforming	✓	✓	✓	Uplink aggregation
[24]	RIS-enabled FL with unified communication–learning design	✓	✓	✓	Uplink aggregation
[25]	Convergence-time optimization for wireless FL	✓			Wireless FL system-level design
This work	RIS-assisted FL dissemination via sum-MI maximization	✓	✓	✓	Downlink model dissemination

RIS-assisted over-the-air FL and proposed a two-step device scheduling and joint beamforming optimization framework that enhanced aggregation reliability and improved learning accuracy under mean-squared error (MSE) constraints. Elbir et al. [23] proposed an FL-based convolutional neural network (CNN) framework for channel estimation in RIS-assisted massive MIMO systems that significantly reduced communication overhead while achieving performance close to centralized learning (CL). Liu et al. [24] further considered a unified communication–learning design for RIS-enabled FL, while Chen et al. [25] studied convergence-time optimization for wireless FL from a system-level perspective.

C. Comparison with Existing Work

The optimization structure of this work shares the standard active and passive beamforming decomposition commonly used in RIS-assisted downlink design, including [18]. However, the problem motivation and system coupling are fundamentally different. Reference [18] optimizes the weighted sum-rate of a generic RIS-aided wireless network, without considering FL model dissemination, round-based training latency, or learning performance. In contrast, our formulation starts from the FL round structure and explicitly models the global-model payload, the downlink completion time, and the feasible number of FL rounds under a fixed time budget. The resulting communication objective is therefore interpreted as an FL-oriented surrogate for reducing model dissemination latency and improving wall-clock learning efficiency.

Table I summarizes the key distinctions between representative existing works and the proposed study. For clarity, the main notations used throughout this paper are summarized in Table II.

TABLE II
MAIN NOTATIONS

Notation	Definition
K	Number of participating devices
N_k	Number of samples at device k
N	Total number of samples
$\mathcal{D}_k$	Local dataset of device k
w	Global model parameter
E	Number of local stochastic gradient descent (SGD) steps
$\mathcal{B}_{k,e}^{(t)}$	Mini-batch of device k at step e in round t
M	Number of BS antennas
R	Number of RIS elements
φ_r	Phase shift of the r-th RIS element
ϕ_r	Reflection coefficient of the r-th RIS element
$\mathbf{\Phi}$	RIS reflection matrix
$\mathbf{h}_{\mathrm{PD},k}$	Direct BS–device channel of device k
$\mathbf{H}_{\mathrm{PR}}$	BS–RIS channel matrix
$\mathbf{h}_{\mathrm{RD},k}$	RIS–device channel of device k
$\mathbf{h}_k(\mathbf{\Phi})$	Effective channel of device k
$\boldsymbol{\theta}$	Set of BS beamforming vectors
γ_k	Signal-to-interference plus noise ratio (SINR) of device k
B	System bandwidth
$R_k(\boldsymbol{\theta}, \mathbf{\Phi})$	Downlink rate of device k
$R_{\mathrm{sum}}(\boldsymbol{\theta}, \mathbf{\Phi})$	System sum MI
$r_k(\boldsymbol{\theta}, \mathbf{\Phi})$	Spectral efficiency of device k
$r_{\min}(\boldsymbol{\theta}, \mathbf{\Phi})$	Minimum user spectral efficiency
$\bar{r}(\boldsymbol{\theta}, \mathbf{\Phi})$	Average user spectral efficiency
$\beta(\boldsymbol{\theta}, \mathbf{\Phi})$	Rate-dispersion factor
S_{dl}	Downlink global-model size
T_{dl}	Downlink dissemination time
$T_{\mathrm{cmp}}^{(t)}$	Computation time in round t
$T_{\mathrm{ul}}^{(t)}$	Uplink transmission time in round t
$T_{\mathrm{round}}^{(t)}$	Wall-clock latency in round t
$\bar{T}_{\mathrm{dl}}, \bar{T}_{\mathrm{cmp}}, \bar{T}_{\mathrm{ul}}$	Average downlink, computation, and uplink latencies
$T_{\max}$	Total training time budget
T	Number of completed global FL rounds
$P_{\max}$	Maximum BS transmit power

III. System Model and Problem Formulation

A. FL Model

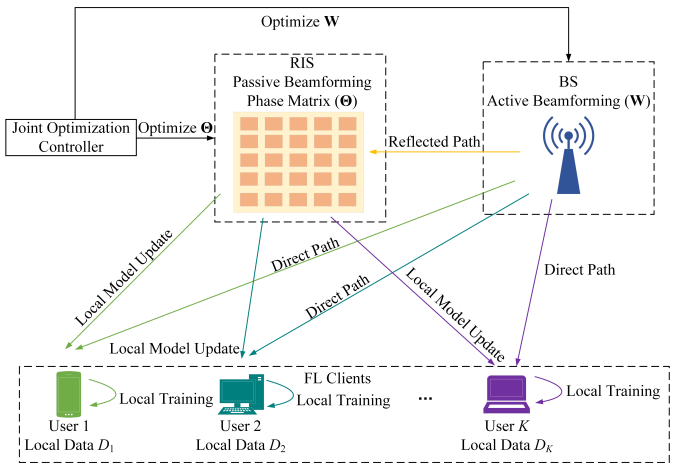

Fig. 1. An FL-assisted RIS communication system model

Fig. 1 presents the architectural layout of the integrated FL and RIS system used in this paper. The primary objective of FL is to minimize the collective training loss. This process can be formally expressed as:

$$\min_{\mathbf{w}\in\mathbb{R}} F(\mathbf{w}) = \sum_{k=1}^{K} \frac{N_k}{N} f_k(\mathbf{w}) \tag{1}$$

where $f_k(\mathbf{w}) = \frac{1}{N_k} \sum_{i\in D_k} L(\mathbf{w}, i)$ represents the loss calculated using model $\mathbf{w}$ over the N_k samples of the k-th device, K is the total number of devices participating in the training, and N is the total number of data samples across all devices. D_k represents the local data index set with cardinality N_k, that is, $N_k = |D_k|$, $L(\mathbf{w}, i)$ is the loss of the i-th sample. In each global FL round, the BS disseminates the updated global model to all devices through parallel downlink transmission. Thus, improving downlink communication quality can help reduce the dissemination latency in the considered setting.

At the beginning of the t-th communication round, the BS disseminates the current global model $\mathbf{w}^{(t)}$ to all participating devices through parallel downlink transmission. Upon receiving $\mathbf{w}^{(t)}$, each device k initializes its local model as

$$\mathbf{w}_k^{(t,0)} = \mathbf{w}^{(t)}. \tag{2}$$

Then, device k performs E steps of local stochastic gradient descent (SGD) on its private dataset D_k. Specifically, at local step $e \in \{0, 1, \ldots, E-1\}$, device k samples a mini-batch $\mathcal{B}_{k,e}^{(t)} \subseteq D_k$ and updates

$$\mathbf{w}_k^{(t,e+1)} = \mathbf{w}_k^{(t,e)} - \eta\, \mathbf{g}_k^{(t,e)}, \tag{3}$$

where η is the learning rate and $\mathbf{g}_k^{(t,e)}$ denotes the mini-batch gradient, which can be defined as

$$\begin{aligned} \mathbf{g}_k^{(t,e)} &\triangleq \nabla_{\mathbf{w}} f_k\left(\mathbf{w}_k^{(t,e)}; \mathcal{B}_{k,e}^{(t)}\right) \\ &= \frac{1}{\left|\mathcal{B}_{k,e}^{(t)}\right|} \sum_{i\in\mathcal{B}_{k,e}^{(t)}} \nabla_{\mathbf{w}} L\left(\mathbf{w}_k^{(t,e)}, i\right), \end{aligned} \tag{4}$$

where i is the sample index and $L(\mathbf{w}, i)$ is the loss on sample i.

After completing E local SGD steps, device k obtains its locally updated model

$$\mathbf{w}_k^{(t+1)} = \mathbf{w}_k^{(t,E)}. \tag{5}$$

Equivalently, device k can form the model difference as

$$\Delta\mathbf{w}_k^{(t)} \triangleq \mathbf{w}_k^{(t+1)} - \mathbf{w}^{(t)}. \tag{6}$$

Each device uploads $\mathbf{w}_k^{(t+1)}$ to the BS, which aggregates all received updates following the FedAvg rule:

$$\mathbf{w}^{(t+1)} = \sum_{k=1}^{K} \frac{N_k}{N} \mathbf{w}_k^{(t+1)} = \mathbf{w}^{(t)} + \sum_{k=1}^{K} \frac{N_k}{N} \Delta\mathbf{w}_k^{(t)}. \tag{7}$$

The above procedure repeats until convergence or a maximum number of communication rounds is reached.

B. RIS Communication Model

Concurrently, the RIS array is deployed to facilitate the communication process. We consider downlink transmission in a system where the BS is equipped with M antennas and serves K single-antenna devices. In this setup, signal propagation occurs via two paths: a direct link from the BS to each device, and an indirect link reflected by the RIS. The RIS comprises R passive reflective elements. Define the reflection coefficient of the r-th reflective element of the RIS as $\varphi_r, r = 1, 2, ..., R$, the reflection coefficient diagonal matrix of the RIS is $\mathbf{\Phi} = \text{diag}(\varphi_1, \varphi_2, ..., \varphi_R)$.

Let $\mathbf{h}_{\text{PD},k} \in \mathbb{C}^{M\times 1}$, $\mathbf{H}_{\text{PR}} \in \mathbb{C}^{M\times R}$, $\mathbf{h}_{\text{RD},k} \in \mathbb{C}^{R\times 1}$ denote the BS–device channel vector, the BS–RIS channel matrix, and the RIS–device channel vector, respectively. For simplicity, the effective channel coefficient vector from the BS to the k-th device can be defined as:

$$\mathbf{h}_k(\mathbf{\Phi}) = \mathbf{h}_{\text{PD},k} + \mathbf{H}_{\text{PR}}\mathbf{\Phi}\mathbf{h}_{\text{RD},k}. \tag{8}$$

In practical RIS-assisted systems, acquiring perfect channel state information (CSI) is challenging because the RIS elements are passive and the cascaded BS–RIS–device channels are high-dimensional. Therefore, the channel coefficients available at the BS are generally estimated from pilot-based channel training. To account for this effect, we write the estimated channels as

$$\hat{\mathbf{h}}_{\text{PD},k} = \mathbf{h}_{\text{PD},k} + \Delta\mathbf{h}_{\text{PD},k} \tag{9}$$

$$\hat{\mathbf{H}}_{\text{PR}} = \mathbf{H}_{\text{PR}} + \Delta\mathbf{H}_{\text{PR}} \tag{10}$$

$$\hat{\mathbf{h}}_{\text{RD},k} = \mathbf{h}_{\text{RD},k} + \Delta\mathbf{h}_{\text{RD},k} \tag{11}$$

where $\Delta\mathbf{h}_{\text{PD},k}$, $\Delta\mathbf{H}_{\text{PR}}$, and $\Delta\mathbf{h}_{\text{RD},k}$ denote the corresponding channel estimation errors. The estimated effective channel used for beamforming design is then given by

$$\hat{\mathbf{h}}_k(\mathbf{\Phi}) = \hat{\mathbf{h}}_{\text{PD},k} + \hat{\mathbf{H}}_{\text{PR}}\mathbf{\Phi}\hat{\mathbf{h}}_{\text{RD},k}. \tag{12}$$

Equivalently, the actual effective channel can be expressed as

$$\mathbf{h}_k(\mathbf{\Phi}) = \hat{\mathbf{h}}_k(\mathbf{\Phi}) + \Delta\mathbf{h}_k(\mathbf{\Phi}), \tag{13}$$

where $\Delta\mathbf{h}_k(\mathbf{\Phi})$ denotes the effective CSI error induced by the direct-link and RIS-cascaded-link estimation errors.

Define the signal transmitted by the BS as $\mathbf{x}$ and the signal received by the k-th device as y_k. The relationship between them satisfies:

$$y_k = \mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi})\,\mathbf{x} + n_k, \tag{14}$$

where $\mathbf{n}_k$ is an additive white Gaussian noise with a mean of 0 and a variance of σ^2.

Suppose the signal component intended for device j is s_j, and it satisfies $\mathbb{E}\left[|s_j|^2\right] = 1$. Here, s_j denotes the user-specific downlink stream that carries an encoded copy of the current global model for device j. Thus, the transmitted signal $\mathbf{x}$ at the BS can be written as

$$\mathbf{x} = \sum_{j=1}^{K} \boldsymbol{\theta}_j s_j, \tag{15}$$

where $\boldsymbol{\theta}_j \in \mathbb{C}^{M\times 1}$ denotes the beamforming vector of the BS. Therefore, the signal received by the k-th device can be rewritten as

$$y_k = \mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi}) \sum_{j=1}^{K} \boldsymbol{\theta}_j s_j + n_k. \tag{16}$$

The above equation can be further decomposed into three parts: the desired signal, the interference signal and the noise, that is,

$$y_k = \mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi})\boldsymbol{\theta}_k s_k + \mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi}) \sum_{j=1, j\neq k}^{K} \boldsymbol{\theta}_j s_j + n_k, \tag{17}$$

where the first part represents the desired received signal for the k-th device, the second part represents the interference caused by streams intended for other devices, and the third part represents the additive noise. Thus, the signal-to-interference-plus-noise ratio for the k-th device can be expressed as

$$\gamma_k = \frac{\left|\left|\mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi})\boldsymbol{\theta}_k\right|\right|^2}{\left|\left|\mathbf{h}_k^{\mathrm{H}}(\mathbf{\Phi}) \sum\limits_{j=1, j\neq k}^{K} \boldsymbol{\theta}_j\right|\right|^2 + \sigma^2}. \tag{18}$$

According to Shannon's capacity formula, the achievable spectral efficiency of device k can be expressed as $\log_2(1+\gamma_k)$. Therefore, the system sum rate is the sum of the MI of all the links (including interference). The sum rate for all user devices can be expressed as

$$R_{\mathrm{sum}} = \sum_{k=1}^{K} \log_2(1+\gamma_k). \tag{19}$$

C. Latency-Aware Formulation via MI Maximization

In each FL round, the BS disseminates the current global model $\mathbf{w}^{(t)}$ to all participating devices before local training starts. Let S_{dl} denote the payload size of the downlink global model per round, which depends on the model dimension and quantization precision. As shown in Fig. 2, each FL round consists of downlink model transmission, local training, and uplink upload; thus, reducing the downlink completion time T_{dl} (dominated by the worst device) is key to enabling more global rounds within a fixed time budget $T_{\max}$.

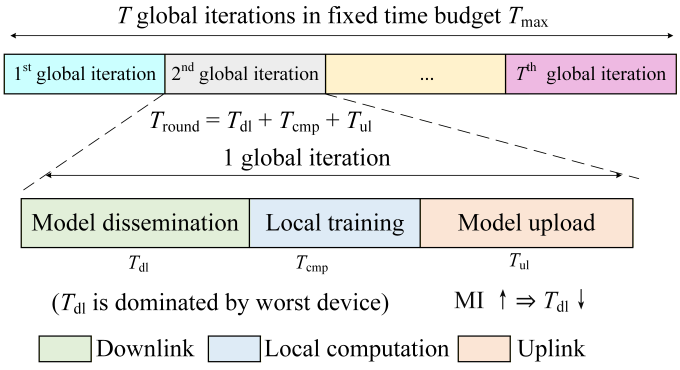

Fig. 2. Round-based timing structure of the RIS-assisted FL system

With the multi-user downlink transmission model in (16)–(18), the achievable spectral efficiency of device k is

$$r_k(\boldsymbol{\theta}, \mathbf{\Phi}) \triangleq \log_2\big(1+\gamma_k(\boldsymbol{\theta}, \mathbf{\Phi})\big), \tag{20}$$

and the corresponding downlink rate is

$$R_k(\boldsymbol{\theta}, \mathbf{\Phi}) = B\, r_k(\boldsymbol{\theta}, \mathbf{\Phi}), \tag{21}$$

where B is the system bandwidth. The system sum MI (sum rate) in (19) can be equivalently written as

$$R_{\mathrm{sum}}(\boldsymbol{\theta}, \mathbf{\Phi}) = \sum_{k=1}^{K} r_k(\boldsymbol{\theta}, \mathbf{\Phi}). \tag{22}$$

Since each device needs to receive S_{dl} bits of the global model in every round, the downlink completion time is governed by the slowest user device in the parallel transmission, i.e.,

$$T_{\mathrm{dl}}(\boldsymbol{\theta}, \mathbf{\Phi}) = \max_{k\in\{1,\ldots,K\}} \frac{S_{\mathrm{dl}}}{R_k(\boldsymbol{\theta}, \mathbf{\Phi})} = \frac{S_{\mathrm{dl}}}{B} \max_k \frac{1}{r_k(\boldsymbol{\theta}, \mathbf{\Phi})}. \tag{23}$$

To further clarify the relationship between the bottleneck downlink latency and the system sum MI, we define the minimum user spectral efficiency and the average spectral efficiency as

$$r_{\min}(\boldsymbol{\theta}, \mathbf{\Phi}) \triangleq \min_{k\in\{1,\ldots,K\}} r_k(\boldsymbol{\theta}, \mathbf{\Phi}), \tag{24}$$

$$\bar{r}(\boldsymbol{\theta}, \mathbf{\Phi}) \triangleq \frac{1}{K} R_{\mathrm{sum}}(\boldsymbol{\theta}, \mathbf{\Phi}). \tag{25}$$

Then, the downlink completion time in (23) can be equivalently written as

$$T_{\mathrm{dl}}(\boldsymbol{\theta}, \mathbf{\Phi}) = \frac{S_{\mathrm{dl}}}{B r_{\min}(\boldsymbol{\theta}, \mathbf{\Phi})}. \tag{26}$$

We further introduce the rate-dispersion factor

$$\beta(\boldsymbol{\theta}, \mathbf{\Phi}) \triangleq \frac{\bar{r}(\theta, \Phi)}{r_{\min}(\boldsymbol{\theta}, \mathbf{\Phi})} = \frac{R_{\mathrm{sum}}(\boldsymbol{\theta}, \mathbf{\Phi})}{K r_{\min}(\boldsymbol{\theta}, \mathbf{\Phi})} \geq 1, \tag{27}$$

which measures the gap between the average user spectral efficiency and the bottleneck user spectral efficiency. Accordingly, the worst-user downlink latency admits the following exact expression:

$$T_{\mathrm{dl}}(\boldsymbol{\theta},\boldsymbol{\Phi}) = \frac{KS_{\mathrm{dl}}}{BR_{\mathrm{sum}}(\boldsymbol{\theta},\boldsymbol{\Phi})}\beta(\boldsymbol{\theta},\boldsymbol{\Phi}). \tag{28}$$

This expression shows that sum-MI maximization is not unconditionally equivalent to worst-user latency minimization. Instead, the two objectives are aligned when the rate-dispersion factor $\beta(\boldsymbol{\theta},\boldsymbol{\Phi})$ is bounded or does not increase significantly after beamforming optimization. Specifically, if the considered FL participant set satisfies

$$\beta(\boldsymbol{\theta},\boldsymbol{\Phi}) \le \beta_{\max}, \tag{29}$$

then the downlink latency is upper-bounded by

$$T_{\mathrm{dl}}(\boldsymbol{\theta},\boldsymbol{\Phi}) \le \frac{K\beta_{\max}S_{\mathrm{dl}}}{BR_{\mathrm{sum}}(\boldsymbol{\theta},\boldsymbol{\Phi})}. \tag{30}$$

Therefore, under bounded rate dispersion, increasing the system sum MI reduces an explicit upper bound on the bottleneck downlink latency.

It should be emphasized that the bounded rate-dispersion condition is not a universal guarantee for arbitrary channel realizations. From (28), for two beamforming designs a and b, we have

$$\frac{T_{\mathrm{dl}}^{(a)}}{T_{\mathrm{dl}}^{(b)}} = \frac{R_{\mathrm{sum}}^{(b)}}{R_{\mathrm{sum}}^{(a)}}\frac{\beta^{(a)}}{\beta^{(b)}}. \tag{31}$$

Therefore, the sum-MI improvement of design a reduces the worst-user downlink latency compared with design b only when

$$\frac{R_{\mathrm{sum}}^{(a)}}{R_{\mathrm{sum}}^{(b)}} > \frac{\beta^{(a)}}{\beta^{(b)}}. \tag{32}$$

This condition clarifies that a large increase in β may offset the benefit of sum-MI maximization. In particular, if some users suffer from severe fading or blockage, $r_{\min}$ may become very small, which leads to a large β and weakens the effectiveness of a pure sum-MI objective for bottleneck latency reduction. In the considered scheduled RIS-assisted FL setting, the simulation results in Section V-D show that β remains moderate, which empirically supports the use of sum-MI maximization as a tractable surrogate.

Directly minimizing (23) is equivalent to maximizing $r_{\min}(\boldsymbol{\theta},\boldsymbol{\Phi})$, i.e., a max-min spectral efficiency problem. This formulation directly captures the FL downlink bottleneck, but it introduces a non-smooth objective and may lead to a conservative beamforming design that sacrifices the aggregate dissemination efficiency of the participating users. Therefore, we do not claim that sum-MI maximization is equivalent to worst-user latency minimization. Instead, based on (28)–(30), sum-MI maximization is adopted as a tractable communication-side surrogate when the rate dispersion among the participating users is bounded.

Besides downlink dissemination, each FL round also includes local computation and uplink reporting. We model the wall-clock latency of round t as

$$T_{\mathrm{round}}^{(t)} = T_{\mathrm{dl}}^{(t)} + T_{\mathrm{cmp}}^{(t)} + T_{\mathrm{ul}}^{(t)}, \tag{33}$$

where $T_{\mathrm{cmp}}^{(t)}$ denotes the local training time, and $T_{\mathrm{ul}}^{(t)}$ denotes the uplink time for transmitting local updates back to the BS. In the simulation part, $T_{\mathrm{cmp}}^{(t)}$ is treated as a fixed constant because all compared schemes use the same local model architecture, number of local SGD steps, batch size, and local computing configuration. Under this homogeneous computation setting, the local computation latency is independent of the downlink beamforming design, and fixing $T_{\mathrm{cmp}}^{(t)}$ helps isolate the contribution of the proposed RIS-assisted downlink optimization. In a heterogeneous deployment, $T_{\mathrm{cmp}}^{(t)}$ can be replaced by a measured or device-dependent computation latency without changing the proposed downlink beamforming algorithm.

If the CSI acquisition overhead is explicitly considered, the round latency can be extended as

$$T_{\mathrm{round}}^{(t)} = T_{\mathrm{CSI}}^{(t)} + T_{\mathrm{dl}}^{(t)} + T_{\mathrm{cmp}}^{(t)} + T_{\mathrm{ul}}^{(t)}, \tag{34}$$

where $T_{\mathrm{CSI}}^{(t)}$ denotes the pilot training and CSI feedback overhead in round t. In this work, $T_{\mathrm{CSI}}^{(t)}$ is not optimized jointly with the beamforming variables. A larger CSI acquisition overhead directly reduces the number of feasible FL rounds under a fixed time budget, while channel estimation errors degrade the realized MI and increase the downlink latency.

In dynamic wireless environments, the channel coefficients and the resulting beamforming variables may vary across FL rounds. Specifically, the direct channel, RIS-assisted channel, and beamforming variables in round t can be written as $\mathbf{h}_{\mathrm{PD},k}^{(t)}$, $\mathbf{H}_{\mathrm{PR}}^{(t)}$, $\mathbf{h}_{\mathrm{RD},k}^{(t)}$, $\boldsymbol{\theta}^{(t)}$, and $\boldsymbol{\Phi}^{(t)}$, respectively. Accordingly, the spectral efficiency and downlink latency become round-dependent, i.e.,

$$r_k^{(t)} = \log_2\left(1+\gamma_k^{(t)}(\boldsymbol{\theta}^{(t)},\boldsymbol{\Phi}^{(t)})\right), \tag{35}$$

and

$$T_{\mathrm{dl}}^{(t)} = \frac{S_{\mathrm{dl}}}{B\min_k r_k^{(t)}}. \tag{36}$$

Therefore, the exact expression of the feasible number of FL rounds does not require a stationary per-round latency; it is determined by the accumulated latency over all completed rounds.

Let $T_{\max}$ be the total time budget for the learning process. The maximum number of communication rounds that can be completed within $T_{\max}$ is

$$T \triangleq \max\left\{\tau\in\mathbb{Z}_+ : \sum_{t=1}^{\tau} T_{\mathrm{round}}^{(t)} \le T_{\max}\right\}. \tag{37}$$

For analytical tractability, we further introduce the empirical average round latency over τ completed rounds:

$$\bar{T}_{\mathrm{round}}(\tau) \triangleq \frac{1}{\tau}\sum_{t=1}^{\tau} T_{\mathrm{round}}^{(t)}. \tag{38}$$

Then, the exact round-count condition can be equivalently expressed as

$$T = \max\left\{\tau\in\mathbb{Z}_+ : \tau\bar{T}_{\mathrm{round}}(\tau) \le T_{\max}\right\}. \tag{39}$$

When the channel and computation processes are quasi-stationary or ergodic over the considered training interval,

$\bar{T}_{\text{round}}(\tau)$ can be approximated by a long-term average latency $\bar{T}_{\text{round}}$, yielding

$$T \approx \left\lfloor \frac{T_{\max}}{\bar{T}_{\text{round}}} \right\rfloor = \left\lfloor \frac{T_{\max}}{\bar{T}_{\text{dl}} + \bar{T}_{\text{cmp}} + \bar{T}_{\text{ul}}} \right\rfloor. \tag{40}$$

This approximation is used only to provide an interpretable link between average downlink latency and the number of feasible FL rounds. In the dynamic case, temporal variations in fading or computation load directly affect $T_{\text{dl}}^{(t)}$ and $T_{\text{cmp}}^{(t)}$, and the exact accumulated-latency expression above should be used.

The ultimate goal of FL is to reduce the training loss within a limited wall-clock time. Under a total time budget $T_{\max}$, one may aim to maximize the feasible number of rounds by optimizing the physical-layer transmission strategy:

$$\begin{aligned} \text{(P0)}: \quad & \max_{\boldsymbol{\theta},\boldsymbol{\Phi}} \quad T \\ & \text{s.t.} \quad \sum_{t=1}^{T} T_{\text{round}}^{(t)} \le T_{\max}, \\ & \qquad \sum_{k=1}^{K} \|\boldsymbol{\theta}_k\|^2 \le P_{\max}, \\ & \qquad \boldsymbol{\Phi} = \text{diag}\left(e^{\text{j}\phi_1}, e^{\text{j}\phi_2}, \ldots, e^{\text{j}\phi_R}\right). \end{aligned}$$

However, (P0) is difficult to handle directly since T is integer-valued and $T_{\text{round}}^{(t)}$ depends on the non-convex rate expressions through γ_k in (18). With the stationary approximation above and treating $\bar{T}_{\text{cmp}}$ and $\bar{T}_{\text{ul}}$ as constants, (P0) can be effectively pursued by improving the downlink transmission efficiency. In implementation, the proposed beamforming design can be applied in a block-wise manner. At the beginning of each FL downlink round, the BS obtains or estimates the current CSI and solves the instantaneous RIS-assisted beamforming problem for the current channel realization. For notational simplicity, we omit the round index t in the following optimization problem and write $(\boldsymbol{\theta}, \boldsymbol{\Phi})$ instead of $(\boldsymbol{\theta}^{(t)}, \boldsymbol{\Phi}^{(t)})$. Motivated by the latency–MI relation in (28)–(30), we adopt sum-MI maximization as a tractable per-round communication-side surrogate for the RIS-assisted FL downlink under bounded rate dispersion, which leads to

$$\begin{aligned} \text{(P1)}: \quad & \max_{\boldsymbol{\theta},\boldsymbol{\Phi}} \quad R_{\text{sum}}(\boldsymbol{\theta}, \boldsymbol{\Phi}) \\ & \text{s.t.} \quad \sum_{k=1}^{K} ||\boldsymbol{\theta}_k||^2 \le P_{\max}, \\ & \qquad \boldsymbol{\Phi} = \text{diag}(\text{e}^{\text{j}\phi_1}, \text{e}^{\text{j}\phi_2}, ..., \text{e}^{\text{j}\phi_R}). \end{aligned} \tag{41}$$

Remark 1*:* Although (P1) has a rate-maximization form similar to conventional RIS-aided downlink designs such as [18], its role here is FL-oriented. The objective is motivated by the round-latency model, where the global-model payload, the downlink completion time, and the total training-time budget jointly determine the number of feasible FL rounds. Thus, R_{sum} is used as a tractable surrogate for improving wall-clock FL efficiency, rather than as a standalone throughput metric.

Remark 2*:* Maximizing $\min_k r_k(\boldsymbol{\theta}, \boldsymbol{\Phi})$ would more directly minimize the bottleneck latency $T_{\text{dl}} = S_{\text{dl}}/(B \min_k r_k)$. However, it leads to a non-smooth and more conservative beamforming problem. In this work, we consider a scheduled FL downlink with bounded rate dispersion, under which maximizing R_{sum} reduces an explicit upper bound on T_{dl} while preserving the tractable WMMSE-based structure. Max-min MI or weighted sum-MI designs can be considered for fairness-critical scenarios.

IV. Alternating Optimization Beamforming

A. Active Beamforming Optimization

For a given RIS phase shift matrix $\boldsymbol{\Phi}$, problem (P1) can be simplified to the optimization of the active beamforming vector $\boldsymbol{\theta}$, i.e.,

$$\begin{aligned} \text{(P2)}: \quad & \max_{\boldsymbol{\theta}} \quad R_{\text{sum}}(\boldsymbol{\theta}) \\ & \text{s.t.} \quad \sum_{k=1}^{K} \|\boldsymbol{\theta}_k\|^2 \le P_{\max}. \end{aligned} \tag{42}$$

Although (P2) is still non-convex, it can be efficiently handled by the classical WMMSE approach. Consider the received signal model in (16)–(18). For each user device k, we introduce a scalar equalizer u_k to estimate the transmitted signal s_k from y_k, yielding

$$\hat{s}_k = u_k^* y_k. \tag{43}$$

Assuming $\mathbb{E}[|s_j|^2] = 1$ and $\mathbb{E}[s_i s_j^*] = 0$ for $i \neq j$, the MSE is defined as

$$E_k \triangleq \mathbb{E}\left[|s_k - \hat{s}_k|^2\right]. \tag{44}$$

By expanding E_k using the signal model, we obtain the following explicit expression

$$\begin{aligned} E_k = |u_k|^2 \left(\sum_{j=1}^{K} \left|\mathbf{h}_k^{\text{H}}(\boldsymbol{\Phi})\boldsymbol{\theta}_j\right|^2 + \sigma^2 \right) \\ - 2\Re\left\{u_k\, \mathbf{h}_k^{\text{H}}(\boldsymbol{\Phi})\boldsymbol{\theta}_k\right\} + 1. \end{aligned} \tag{45}$$

A well-known rate–MSE identity shows that the achievable rate $\log_2(1+\gamma_k)$ is equivalently expressed as

$$\log_2(1+\gamma_k) = \max_{u_k,\, w_k>0} \left\{ \log_2(w_k) - \frac{w_k E_k}{\ln 2} + \frac{1}{\ln 2} \right\}, \tag{46}$$

where w_k is an auxiliary weight variable. At the optimum, u_k is the minimum mean-square error (MMSE) equalizer and $w_k = 1/E_k$. Therefore, maximizing the sum rate is equivalent to minimizing a weighted sum MSE objective, which motivates an AO over $\{u_k, w_k, \boldsymbol{\theta}_k\}$.

We solve (P2) by iteratively optimizing the equalizer u_k, weight w_k, and beamforming vector $\boldsymbol{\theta}_k$.

(1) **Update Equalizer u_k:** For a given $\boldsymbol{\theta}$ and $\boldsymbol{\Phi}$, minimizing E_k w.r.t. u_k yields the optimal MMSE equalizer

$$u_k = \left(\sum_{j=1}^{K} \left|\mathbf{h}_k^{\text{H}}(\boldsymbol{\Phi})\boldsymbol{\theta}_j\right|^2 + \sigma^2 \right)^{-1} \mathbf{h}_k^{\text{H}}(\boldsymbol{\Phi})\boldsymbol{\theta}_k. \tag{47}$$

(2) **Update Weight w_k:** With u_k fixed, maximizing (46) over $w_k > 0$ gives

$$w_k = \frac{1}{E_k}, \tag{48}$$

where E_k is computed by (45). Substituting (47) into (45) gives the minimum MSE, and (48) follows the standard WMMSE rule.

(3) **Update Beamforming Vectors $\boldsymbol{\theta}_k$:** For given sets $\{u_k, w_k\}$ and fixed $\boldsymbol{\Phi}$, maximizing the sum rate is equivalent to minimizing the weighted sum of MSEs:

$$\begin{aligned} \min_{\boldsymbol{\theta}} \quad & \sum_{k=1}^{K} w_k E_k(\boldsymbol{\theta}, \boldsymbol{\Phi}, u_k) \\ \text{s.t.} \quad & \sum_{k=1}^{K} \|\boldsymbol{\theta}_k\|^2 \le P_{\max}. \end{aligned} \tag{49}$$

By substituting (45) into (49) and collecting terms, the problem becomes a convex quadratically constrained quadratic program (QCQP). Define

$$\mathbf{A} \triangleq \sum_{j=1}^{K} w_j |u_j|^2 \, \mathbf{h}_j(\boldsymbol{\Phi}) \mathbf{h}_j^{\mathrm{H}}(\boldsymbol{\Phi}), \tag{50}$$

$$\mathbf{b}_k \triangleq w_k u_k^* \, \mathbf{h}_k(\boldsymbol{\Phi}). \tag{51}$$

Then the Karush-Kuhn-Tucker (KKT) optimality condition yields the closed-form structure

$$\boldsymbol{\theta}_k = (\mathbf{A} + \lambda \mathbf{I}_M)^{-1} \mathbf{b}_k, \tag{52}$$

where $\lambda \ge 0$ is the Lagrange multiplier associated with the transmit power constraint. If $\sum_{k=1}^{K} \|\boldsymbol{\theta}_k\|^2 < P_{\max}$, we have $\lambda = 0$; otherwise, $\lambda > 0$ is chosen such that $\sum_{k=1}^{K} \|\boldsymbol{\theta}_k\|^2 = P_{\max}$. Since $\sum_{k=1}^{K} \|\boldsymbol{\theta}_k(\lambda)\|^2$ is monotonically decreasing in λ, λ can be efficiently found by a bisection search.

Remark 3*:* (52) ensures the dimensions are consistent, since $\mathbf{h}_k(\boldsymbol{\Phi}) \in \mathbb{C}^{M\times 1}$ and thus $\boldsymbol{\theta}_k \in \mathbb{C}^{M\times 1}$.

B. Passive Beamforming Optimization

When the active beamforming vectors $\boldsymbol{\theta}$ are fixed, the RIS phase shift matrix $\boldsymbol{\Phi}$ can be optimized as

$$\begin{aligned} \text{(P3):} \quad \max_{\boldsymbol{\Phi}} \quad & R_{\text{sum}}(\boldsymbol{\Phi}) \\ \text{subject to} \quad & |\varphi_r| = 1, \forall r. \end{aligned} \tag{53}$$

Problem (P3) remains non-convex due to the unit-modulus constraint.

To simplify notation, define the RIS phase vector

$$\mathbf{v} \triangleq \left[e^{\mathrm{j}\varphi_1}, e^{\mathrm{j}\varphi_2}, \ldots, e^{\mathrm{j}\varphi_R}\right]^{\mathrm{T}} \in \mathbb{C}^{R\times 1}, \tag{54}$$

$$\boldsymbol{\Phi} = \mathrm{diag}(\mathbf{v}). \tag{55}$$

For user device k and stream j, the effective channel term can be rewritten as

$$\mathbf{h}_k^{\mathrm{H}}(\boldsymbol{\Phi})\boldsymbol{\theta}_j = \mathbf{h}_{\mathrm{PD},k}^{\mathrm{H}}\boldsymbol{\theta}_j + \mathbf{h}_{\mathrm{RD},k}^{\mathrm{H}}\boldsymbol{\Phi}\mathbf{H}_{\mathrm{PR}}^{\mathrm{H}}\boldsymbol{\theta}_j = c_{k,j} + \mathbf{v}^{\mathrm{H}}\mathbf{b}_{k,j}, \tag{56}$$

where

$$c_{k,j} \triangleq \mathbf{h}_{\mathrm{PD},k}^{\mathrm{H}}\boldsymbol{\theta}_j, \tag{57}$$

$$\mathbf{b}_{k,j} \triangleq \mathrm{diag}\big(\mathbf{h}_{\mathrm{RD},k}^{*}\big)\, \mathbf{H}_{\mathrm{PR}}^{\mathrm{H}}\boldsymbol{\theta}_j \in \mathbb{C}^{R\times 1}. \tag{58}$$

Algorithm 1 Joint Active and Passive Beamforming for RIS-Assisted FL Downlink

Require: Channel state information $\{\mathbf{h}_{\mathrm{PD},k}, \mathbf{h}_{\mathrm{RD},k}\}_{k=1}^{K}$ and $\mathbf{H}_{\mathrm{PR}}$; power budget $P_{\max}$; maximum outer iterations I_{AO}; maximum inner iterations I_{W} and I_{B}; tolerances $\epsilon_{\mathrm{AO}}, \epsilon_{\mathrm{WMMSE}}, \epsilon_{\mathrm{BCD}}$.

Ensure: BS Active Beamforming vector $\boldsymbol{\theta}$, RIS phase matrix $\boldsymbol{\Phi}$, RIS phase vector $\mathbf{v}$.

Initialization: choose $\mathbf{v}^{(0)}$ with $|v_r^{(0)}| = 1$; set $\boldsymbol{\Phi}^{(0)} = \mathrm{diag}(\mathbf{v}^{(0)})$; set $t \leftarrow 0$.
Initialize $\boldsymbol{\theta}^{(0)}$ and compute $R_{\text{sum}}^{(0)}$.
repeat
1) Active beamforming update (WMMSE at BS): fix $\boldsymbol{\Phi}^{(t)}$.
  Set $\boldsymbol{\theta} \leftarrow \boldsymbol{\theta}^{(t)}$.
  for $i = 1$ to I_{W} **do**
    Update equalizers u_k according to (47).
    Update weights w_k by (48).
    Update beamforming vector $\boldsymbol{\theta}_k$ using (52).
    Find λ by bisection to satisfy power constraint.
    Stopping (inner): break if the relative increase of R_{sum} is below $\epsilon_{\mathrm{WMMSE}}$.
  end for
  Set $\boldsymbol{\theta}^{(t+1)} \leftarrow \boldsymbol{\theta}$.
2) Passive beamforming update (BCD at RIS): fix $\boldsymbol{\theta}^{(t+1)}$.
  Compute auxiliary quantities $c_{k,j}, \mathbf{b}_{k,j}$ and form $(\mathbf{Q}, \mathbf{p})$ in (60)–(61).
  Set $\mathbf{v} \leftarrow \mathbf{v}^{(t)}$.
  for $i = 1$ to I_{B} **do**
    for $r = 1$ to R **do**
      Update φ_r and $v_r = e^{\mathrm{j}\varphi_r}$ using (63).
    end for
    Stopping (inner): break if the objective decrease is below ϵ_{BCD}.
  end for
  Set $\mathbf{v}^{(t+1)} \leftarrow \mathbf{v}$ and $\boldsymbol{\Phi}^{(t+1)} \leftarrow \mathrm{diag}(\mathbf{v}^{(t+1)})$.
  Compute $R_{\text{sum}}^{(t+1)}$ with $(\boldsymbol{\theta}^{(t+1)}, \boldsymbol{\Phi}^{(t+1)})$.
  $t \leftarrow t + 1$.
until $t \ge I_{\mathrm{AO}}$ **or** $\frac{|R_{\text{sum}}^{(t)} - R_{\text{sum}}^{(t-1)}|}{\max\{1, R_{\text{sum}}^{(t-1)}\}} \le \epsilon_{\mathrm{AO}}$.
return $(\boldsymbol{\theta}^{(t)}, \boldsymbol{\Phi}^{(t)})$.

Substituting (56) into the WMMSE objective with u_k, w_k fixed yields a quadratic function of $\mathbf{v}$:

$$\min_{|v_r|=1, \forall r} \quad \mathbf{v}^{\mathrm{H}}\mathbf{Q}\mathbf{v} + 2\Re\{\mathbf{v}^{\mathrm{H}}\mathbf{p}\} + C, \tag{59}$$

where C is a constant, and $\mathbf{Q}$, $\mathbf{p}$ are given by

$$\mathbf{Q} = \sum_{k=1}^{K} w_k |u_k|^2 \sum_{j=1}^{K} \mathbf{b}_{k,j}\mathbf{b}_{k,j}^{\mathrm{H}}, \tag{60}$$

$$\mathbf{p} = \sum_{k=1}^{K} w_k \left(\sum_{j\neq k} |u_k|^2 c_{k,j}\mathbf{b}_{k,j} - (1 - u_k^* c_{k,k}) u_k \, \mathbf{b}_{k,k} \right). \tag{61}$$

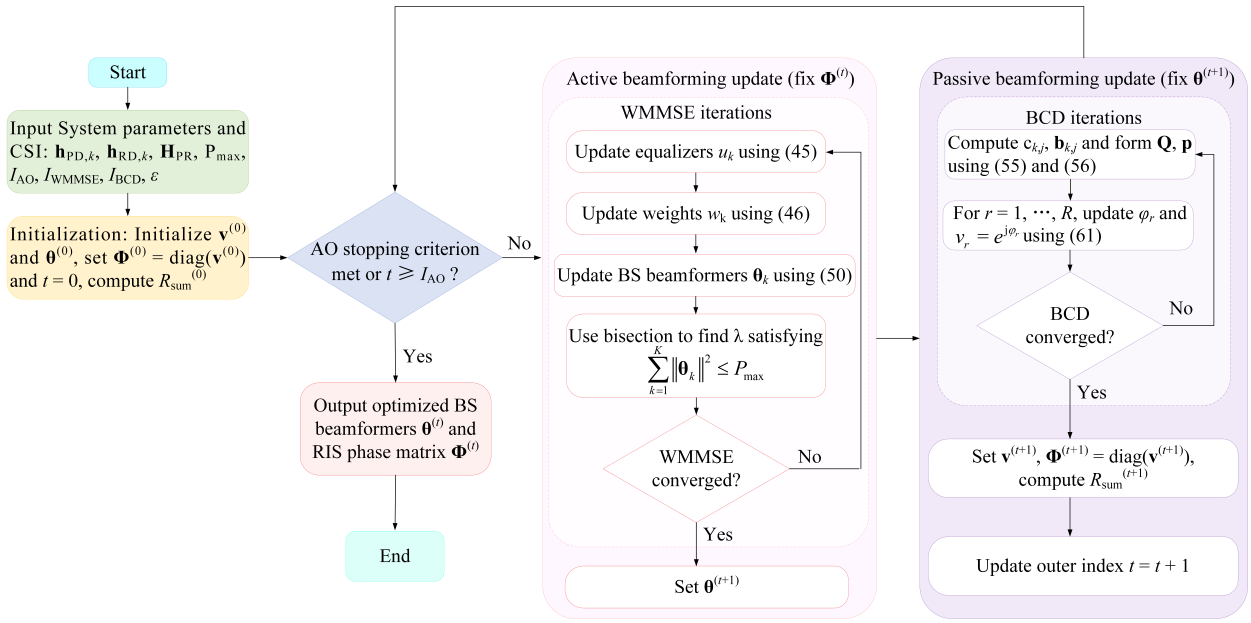

Fig. 3. Flowchart of the proposed AO beamforming algorithm

Problem (59) can be efficiently solved using BCD. Specifically, each element $v_r = e^{j\varphi_r}$ of $\mathbf{v}$ is updated while keeping the others fixed. For a given r, the objective can be written as

$$f(v_r) = q_{r,r}|v_r|^2 + 2\Re\left\{ v_r^* \left(\sum_{i \neq r} q_{i,r} v_i + p_r \right) \right\} + C, \quad (62)$$

where $q_{i,r}$ is the (i,r)-th element of $\mathbf{Q}$ and p_r is the r-th element of $\mathbf{p}$. Under the unit-modulus constraint $|v_r| = 1$, minimizing (62) is achieved by setting the phase of v_r opposite to the phase of the linear term, i.e.,

$$\varphi_r = \arg\left(-\sum_{i \neq r} q_{i,r} v_i - p_r \right), \qquad v_r \leftarrow e^{j\varphi_r}. \quad (63)$$

The proposed AO beamforming algorithm is summarized in Algorithm 1, while the flowchart is shown in Fig. 3.

Algorithm 1 is implemented using the CSI available at the BS. Under perfect CSI, the estimated channel is identical to the actual channel. Under imperfect CSI, the beamforming variables are optimized based on $\{\hat{\mathbf{h}}_k(\boldsymbol{\Phi})\}_{k=1}^K$, while the actually achieved MI is determined by $\{\mathbf{h}_k(\boldsymbol{\Phi})\}_{k=1}^K$. Therefore, the monotonic improvement property holds for the estimated sum-MI objective, but the realized sum MI may degrade when the CSI error increases.

The proposed AO framework can still be applied under imperfect CSI by replacing $\mathbf{h}_k(\boldsymbol{\Phi})$ with $\hat{\mathbf{h}}_k(\boldsymbol{\Phi})$ in the WMMSE and BCD updates. The resulting design is computationally unchanged, but its performance depends on the CSI quality. If robust performance is required, the current formulation can be extended to a worst-case robust design, that is,

$$\max_{\boldsymbol{\theta},\boldsymbol{\Phi}} \min_{\{\Delta\mathbf{h}_k : \|\Delta\mathbf{h}_k\| \leq \varepsilon_k\}} R_{\mathrm{sum}}\big(\boldsymbol{\theta}, \boldsymbol{\Phi}; \{\hat{\mathbf{h}}_k + \Delta\mathbf{h}_k\}_{k=1}^K\big). \quad (64)$$

This robust extension can be handled by conservative approximation, but it introduces additional computational complexity. Since the main focus of this paper is the latency-aware RIS-assisted FL downlink design, we keep the main algorithm based on estimated CSI.

C. Convergence and Complexity Analysis

By construction, each update step in Algorithm 1 does not increase the WMMSE objective for the corresponding block variables. Through the rate–WMMSE equivalence in (46), the system sum-rate R_{sum} is non-decreasing across outer iterations. Since R_{sum} is upper-bounded under the transmit power constraint, the alternating procedure converges to a stationary point of (P1).

The per-iteration computational complexity is dominated by: i) the active beamforming update, which requires forming $\mathbf{A}$ in (50) and solving (52), involving an $M \times M$ matrix inversion and a bisection search over λ; and ii) the passive beamforming update, which updates R phase elements via BCD. In practice, the overall complexity can be summarized as

$$\mathcal{O}\left(I_{\mathrm{AO}} \left[I_{\mathrm{W}} \left(I_{\mathrm{bis}} M^3 + KM^2 + K^2 M \right) + K^2 R^2 + I_{\mathrm{B}} R^2 \right] \right), \quad (65)$$

where I_{AO} is the number of outer AO iterations, I_{W} is the number of inner WMMSE iterations, I_{bis} is the number of bisection steps for finding the Lagrange multiplier, and I_{B} is the number of BCD sweeps for RIS phase optimization. The term $I_{\mathrm{W}}(I_{\mathrm{bis}}M^3 + KM^2 + K^2M)$ comes from the WMMSE-based active beamforming update, including the matrix inversion, bisection search, and multi-user interference-term computation. The term K^2R^2 is due to the construction of $\mathbf{Q}$ and $\mathbf{p}$ in (60)-(61), while $I_{\mathrm{B}}R^2$ corresponds to the element-wise BCD update of the RIS phase vector.

The parameters I_{W} and I_{B} specify the maximum numbers of inner iterations for the active and passive beamforming subproblems, respectively, while $\epsilon_{\mathrm{WMMSE}}$, ϵ_{BCD}, and ϵ_{AO}

control the corresponding early stopping conditions. These parameters mainly determine the tradeoff between solution accuracy and computational cost. A larger I_{W} or I_{B} allows the corresponding subproblem to be solved more accurately, which may improve the achieved sum MI in the early iterations, but also increases the computational time almost linearly according to the complexity expression in (65). Similarly, smaller stopping thresholds lead to more accurate block updates and tighter convergence, but may bring only marginal objective improvement after the algorithm is close to a stationary point. In practice, the inner loops usually converge within a small number of iterations because each WMMSE or BCD update monotonically improves the corresponding block objective. Therefore, setting very large I_{W} and I_{B} is unnecessary. The maximum iteration numbers are used only as safeguards, while the relative-improvement thresholds provide adaptive termination.

TABLE III
COMPLEXITY COMPARISON WITH REPRESENTATIVE RIS BEAMFORMING ALGORITHMS

Algorithm	Dominant computational complexity
Semidefinite relaxation (SDR)-based AO [26]	$\mathcal{O}\big(I_{\mathrm{AO}}\left(\mathcal{C}_{\mathrm{ABF}}+I_{\mathrm{SDR}}(R+1)^{6.5}\right)\big)$
Unified manifold optimization (UMO) method [27]	$\mathcal{O}\big(I_{\mathrm{AO}}\left(\mathcal{C}_{\mathrm{ABF}}+I_{\mathrm{RCG}}K^2R^2\right)\big)$
Fractional programming (FP)-based method [18]	$\mathcal{O}\big(I_{\mathrm{FP}}\left(K^2R^2+KMR+KM^2\right)\big)$
Proposed method	$\mathcal{O}\big(I_{\mathrm{AO}}\left(\mathcal{C}_{\mathrm{ABF}}+K^2R^2+I_{\mathrm{B}}R^2\right)\big)$

To further demonstrate the computational efficiency of the proposed algorithm, Table III compares its dominant complexity with representative RIS beamforming algorithms, including SDR-based AO, UMO, and FP-based weighted sum rate maximization. For compactness, define $\mathcal{C}_{\mathrm{ABF}} = I_{\mathrm{W}}\left(I_{\mathrm{bis}}M^3+KM^2+K^2M\right)$. In Table III, I_{SDR}, I_{RCG}, and I_{FP} denote the corresponding numbers of iterations of the SDR solver, Riemannian conjugate gradient (RCG) procedure, and FP-based algorithm, respectively.

As shown in Table III, the proposed method avoids the high-order semidefinite programming operation required by SDR-based AO methods and does not require manifold line search as in UMO methods. Its passive beamforming update only involves forming the quadratic coefficients and performing element-wise BCD updates. Compared with the FP-based weighted sum rate method in [18], the proposed method remains in the same low-complexity BCD family, while it is specifically developed for the FL-oriented downlink dissemination problem. Therefore, the proposed algorithm provides a practical balance between computational efficiency and FL-oriented communication performance.

V. SIMULATION RESULTS

A. Simulation Settings

We evaluate the proposed RIS-assisted FL downlink design under a multi-user multiple-input single-output dissemination setting, where the BS is equipped with M antennas and serves K single-antenna devices with the aid of an R-element RIS. The wireless channels are generated following a Ricean fading model, which captures the coexistence of a dominant specular component and a scattered component and is widely adopted to reflect practical propagation conditions in RIS-assisted links. In the considered geometry, the BS, RIS, and devices are placed at fixed coordinates, as summarized in Table IV, so that both the direct BS–device path and the RIS-reflected path contribute to the effective downlink channel.

To ensure a fair comparison from the learning perspective, all communication-learning benchmark pipelines employ the same neural network architecture and training hyperparameters. Specifically, we adopt a CNN consisting of two 5×5 convolutional layers, each followed by 2×2 max pooling, and then normalization, fully connected layers, ReLU activations, and a Softmax output layer. The learning rate is set to 0.05, and the training process runs for 500 global rounds. Unless otherwise specified, the default communication setting uses signal-to-noise ratio (SNR) $=20$ dB, $K=40$ devices, $M=4$ BS antennas, and $R=64$ RIS reflecting elements.

The simulations use a block-fading setting where the channel remains fixed within one downlink transmission block and can be re-estimated before the next beamforming update. This setting isolates the impact of the proposed RIS-assisted beamforming design, while the theoretical round-latency model in Section III-C also supports time-varying round latencies through $T_{\mathrm{round}}^{(t)}$. Unless otherwise specified, all communication-side results are obtained by Monte Carlo averaging over 20 independent random channel realizations. In the simulations, the maximum number of outer AO iterations is set to $I_{\mathrm{AO}}=50$. The outer AO loop is terminated when the relative improvement of the sum MI satisfies $\frac{|R_{\mathrm{sum}}^{(t)}-R_{\mathrm{sum}}^{(t-1)}|}{\max\{1,R_{\mathrm{sum}}^{(t-1)}\}}\le\epsilon_{\mathrm{AO}}$, where $\epsilon_{\mathrm{AO}}=10^{-4}$ is used unless otherwise specified. The inner WMMSE and BCD stopping thresholds are set to $\epsilon_{\mathrm{WMMSE}}=10^{-4}$ and $\epsilon_{\mathrm{BCD}}=10^{-4}$, respectively.

Different from conventional RIS sum-rate studies, the evaluation includes not only communication-side metrics, such as sum MI and user MI, but also FL-oriented metrics, including downlink dissemination latency and learning performance under standard image classification tasks. This is intended to verify whether the communication gain can translate into improved wall-clock FL progress. In each global round, the BS disseminates the current global model using the beamforming design produced by the corresponding method. Therefore, an increased MI is empirically associated with a shorter downlink dissemination time under the considered setup. We assume perfect channel state information at the BS for both direct and RIS-reflected links, following standard practice in the RIS literature, in order to isolate the gains brought by joint active and passive beamforming optimization. The compared methods differ in both the FL aggregation pipeline and the adopted downlink transmission design; therefore, the comparisons are intended to evaluate end-to-end system behavior rather than isolate a single beamforming component.

TABLE IV
SIMULATION PARAMETERS

Parameters	Values
Coordinate of BS	(0, -50m, 3m)
Coordinate of RIS	(20m, 10m, 5m)
Coordinate of devices	(40m, 0, 0)
Learning rate	0.05
SNR	20dB
Number of devices	40
Number of BS antennas	4
Number of RIS reflecting elements	64
Number of training rounds	500

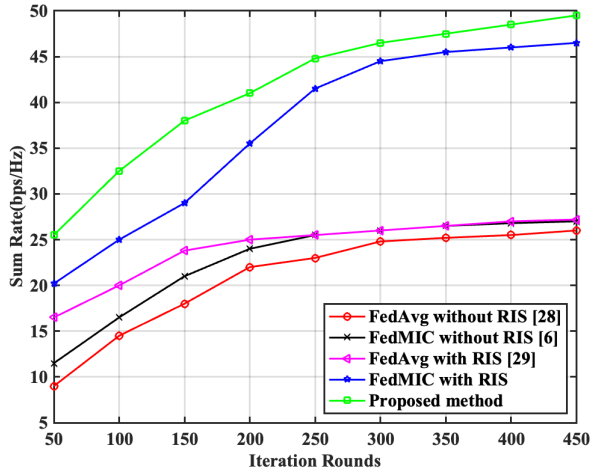

Fig. 4. Comparison of system sum rate for different methods with MNIST dataset

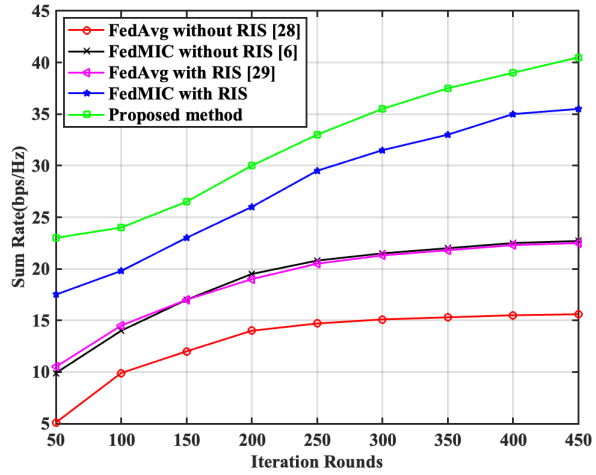

Fig. 5. Comparison of system sum rate for different methods with CIFAR-10 dataset

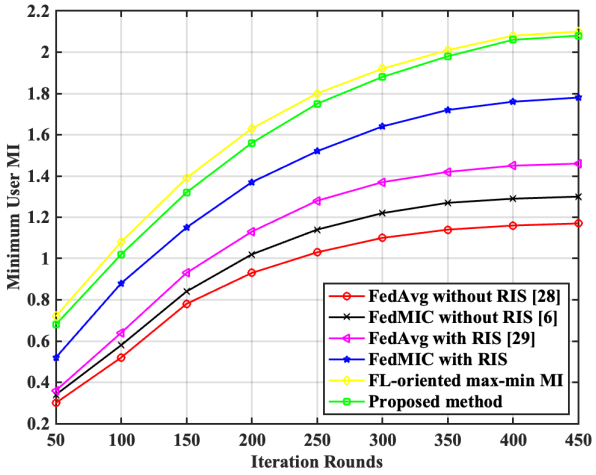

Fig. 6. Comparison of minimum user MI for different methods with MNIST dataset

B. Simulation on System Sum Rate

Firstly, we compare the system sum rate achieved by different transmission designs on both the MNIST and CIFAR-10 datasets (Figs. 4-5). The curves clearly demonstrate the benefit of introducing an RIS into the downlink: compared with the "without RIS" baselines, RIS-assisted schemes consistently achieve higher sum rates due to the additional controllable reflected path that effectively reshapes the propagation environment.

More importantly, among all RIS-assisted approaches, the proposed beamforming strategy delivers the best performance throughout the iterative procedure. This indicates that optimizing the BS beamforming vector alone is insufficient to fully exploit the available degrees of freedom when an RIS is deployed; instead, a coordinated design that jointly adapts the BS beamforming vectors and RIS phases is needed to maximize the aggregate MI. As depicted in the figures, the sum rate achieved by the proposed method can be nearly twice that of the FedAvg baseline without RIS, which highlights the effectiveness of the proposed optimization framework in strengthening the downlink model dissemination links.

C. Simulation on User MI

Meanwhile, we further examine the user-level MI to assess how the proposed design affects the bottleneck downlink quality. To address the FL-specific downlink comparison, we further include an FL-oriented max-min MI baseline, which is inspired by [30] and optimized for our framework. Figs. 6–7 plot the minimum user MI, i.e., the MI of the worst downlink user, versus the number of communication rounds on MNIST and CIFAR-10, respectively. This metric is particularly relevant to FL downlink dissemination, since all participating devices must receive the global model before local training starts, and thus the round latency is governed by the slowest user.

Two consistent observations can be made. First, RIS deployment improves the minimum user MI compared with the non-RIS counterparts, confirming that the controllable reflection path can enhance not only the aggregate link quality but also the bottleneck user link. Second, the proposed joint

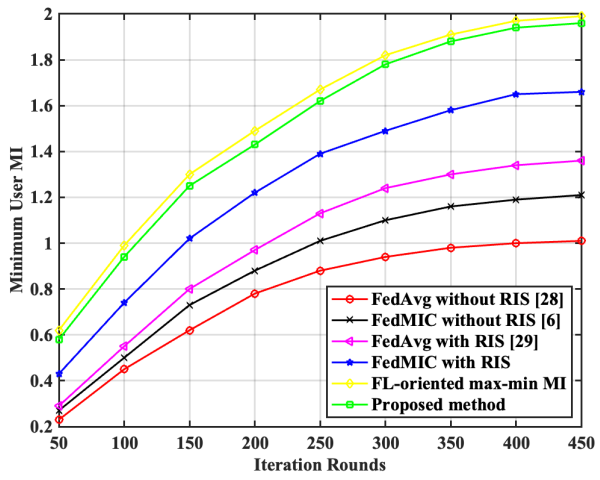

Fig. 7. Comparison of minimum user MI for different methods with CIFAR-10 dataset

beamforming design achieves competitive minimum user MI compared with the FL-oriented max-min MI baseline, while maintaining its advantage in aggregate MI performance.

D. Simulation of the Rate-Dispersion Factor

To empirically verify the bounded rate-dispersion assumption used in (28)–(30), we further evaluate the rate-dispersion factor β under the same Monte Carlo setting. Table V reports the Monte Carlo statistics of the rate-dispersion factor β for the proposed method. For each channel realization, β is computed as the ratio between the average user spectral efficiency and the minimum user spectral efficiency after joint active and passive beamforming optimization. The results show that β remains within a bounded range under all considered settings. Increasing the number of RIS elements generally does not cause a noticeable increase in β, indicating that the sum-MI gain is not obtained by severely sacrificing the bottleneck user. When the number of users increases, β may slightly increase because the probability of having a weaker bottleneck user becomes higher. Nevertheless, the variation remains moderate, which empirically supports the bounded rate-dispersion assumption used in (28)–(30).

E. Simulation on Round Latency

Furthermore, we evaluate the latency performance to quantify how the proposed joint active and passive beamforming accelerates RIS-assisted FL from a wall-clock perspective. We vary one of three system parameters while fixing the other two. Specifically, the downlink dissemination latency T_{dl} reflects the time required for delivering the global model to all devices in each communication round. To isolate the impact of downlink beamforming and RIS phase optimization on the global model dissemination latency, we set the local computation time and uplink reporting time to fixed values across all compared schemes. This is because the considered baselines differ only in the downlink transmission design, while they share identical local training hyperparameters and uplink access strategy. We further note that in practical heterogeneous deployments, T_{cmp} and T_{ul} can be quantified by measuring per-round local training time and by modeling the uplink transmission time based on the update size and uplink spectral efficiency, respectively.

As shown in Table VI, since the proposed design explicitly enhances the effective channels through coordinated BS beamforming and RIS phase configuration, it improves users' achievable rates, which is consistent with the observed reduction in downlink completion time. This reduction is particularly pronounced in challenging propagation conditions, where the RIS contribution becomes dominant. Moreover, even when computation and uplink overheads are fixed, shortening T_{dl} directly lowers T_{round}, which implies that more FL rounds can be completed within a given time budget, thereby improving learning progress per unit time. The latency comparison therefore complements the sum rate plots by demonstrating the practical training speed advantage of the proposed communication design under realistic system constraints.

TABLE V
STATISTICS OF THE RATE-DISPERSION FACTOR β UNDER DIFFERENT SYSTEM SETTINGS

Scenario	Setting	Mean of β	Std. of β
RIS elements	$R = 16$	1.58	0.16
	$R = 32$	1.51	0.14
	$R = 64$	1.47	0.13
	$R = 128$	1.40	0.10
Number of users	$K = 4$	1.08	0.07
	$K = 8$	1.14	0.06
	$K = 12$	1.17	0.06
	$K = 16$	1.21	0.08
SNR	0 dB	1.35	0.13
	5 dB	1.40	0.12
	10 dB	1.43	0.11
	20 dB	1.47	0.13

TABLE VI
DOWNLINK LATENCY COMPARISON UNDER DIFFERENT SYSTEM SETTINGS

Scenario	No RIS baseline	Random RIS phases	Zero-forcing	Proposed
Latency vs. RIS elements R (fixed $K = 40$, SNR= 20 dB)				
$R = 16$	0.120	0.110	0.095	**0.080**
$R = 32$	0.120	0.102	0.087	**0.072**
$R = 64$	0.120	0.095	0.080	**0.065**
$R = 128$	0.120	0.088	0.074	**0.058**
Latency vs. number of users K (fixed $R = 64$, SNR= 20 dB)				
$K = 4$	0.030	0.024	0.020	**0.016**
$K = 8$	0.044	0.038	0.031	**0.025**
$K = 12$	0.058	0.046	0.043	**0.036**
$K = 16$	0.071	0.059	0.053	**0.044**
Latency vs. SNR (dB) (fixed $R = 64$, $K = 40$)				
SNR= 0	0.450	0.360	0.310	**0.260**
SNR= 5	0.280	0.220	0.190	**0.160**
SNR= 10	0.180	0.140	0.120	**0.100**
SNR= 20	0.120	0.095	0.080	**0.065**

TABLE VII
COMMUNICATION–COMPUTATION EFFICIENCY UNDER DIFFERENT LOCAL COMPUTATION LOADS

T_{cmp} (s)	Completed FL rounds within $T_{\max} = 10$ s				Speedup
	No RIS	Random RIS	Zero-forcing	Proposed	
0.02	55	64	71	**80**	1.45×
0.05	47	54	58	**64**	1.36×
0.10	38	42	45	**48**	1.26×
0.20	27	29	31	**32**	1.19×

To further validate the joint communication–computation characteristics of the considered FL system, we evaluate the wall-clock training efficiency under different local computation loads. For each method, the per-round latency is computed as $T_{\text{round}} = T_{\text{dl}} + T_{\text{cmp}} + T_{\text{ul}}$. Under a fixed total training-time budget $T_{\max}$, the number of feasible global rounds is given by (40). As shown in Table VII, the proposed method enables more FL rounds to be completed within the same wall-clock time budget by reducing the downlink dissemination latency. The relative gain is more pronounced when the local computation time is small, indicating that the proposed communication design is particularly beneficial in communication-limited FL scenarios. When T_{cmp} becomes dominant, the speedup naturally decreases because the downlink latency accounts for a smaller fraction of the total round latency.

F. Simulation on Classification Performance

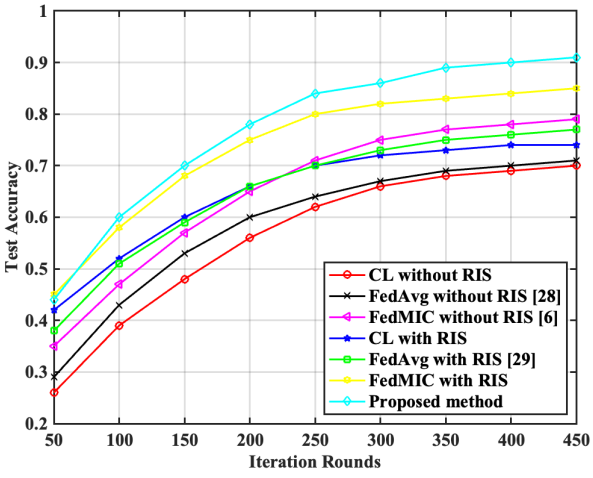

Fig. 8. Comparison of test accuracy for different methods with MNIST dataset

Finally, we compare several performance metrics of the model for the image classification task, including test accuracy, precision, recall, and F1-score. We also include a CL benchmark, in which users upload their local datasets to the BS for centralized training. The comparison of test accuracy is shown in Figs. 8-9, while the other metrics are presented in Table VIII.

The results depicted in Figs. 8 and 9 demonstrate the effectiveness of the proposed design under the considered setup. By improving the wireless communication conditions for FL, the proposed method is empirically associated with faster training progress and higher final test accuracy. The consistent advantage observed across both datasets suggests that enhanced communication rate and reliability can serve as a promising communication-side mechanism for improving the learning performance of wireless FL systems.

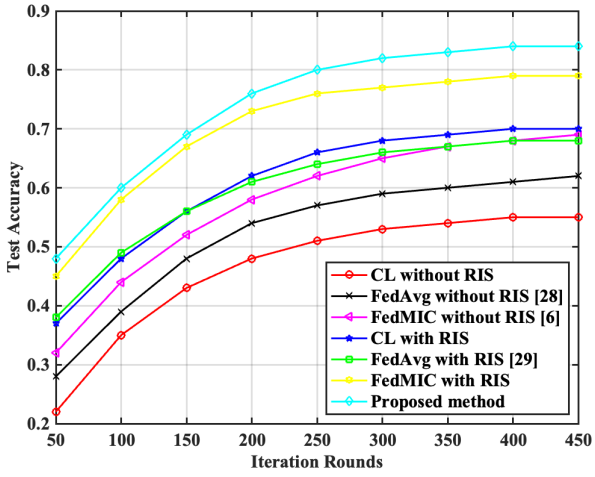

Fig. 9. Comparison of test accuracy for different methods with CIFAR-10 dataset

The empirical results in Table VIII further show that the proposed framework achieves favorable precision, recall, and F1-score on both datasets. These results indicate that the proposed communication-aware beamforming design provides consistent end-to-end performance gains over the compared methods.

TABLE VIII
COMPARISON OF PERFORMANCE METRICS FOR DIFFERENT METHODS AND DATASETS

Method	Dataset	Precision	Recall	F1-score
CL without RIS	MNIST	0.67	0.80	0.73
	CIFAR-10	0.55	0.54	0.54
FedAvg without RIS	MNIST	0.71	0.70	0.70
	CIFAR-10	0.61	0.60	0.60
FedMIC without RIS	MNIST	0.78	0.80	0.79
	CIFAR-10	0.67	0.70	0.68
CL with RIS	MNIST	0.76	0.70	0.73
	CIFAR-10	0.69	0.72	0.70
FedAvg with RIS	MNIST	0.78	0.76	0.77
	CIFAR-10	0.67	0.70	0.68
FedMIC with RIS	MNIST	0.84	0.86	0.85
	CIFAR-10	0.79	0.76	0.77
Proposed method	MNIST	**0.88**	**0.92**	**0.90**
	CIFAR-10	**0.81**	**0.86**	**0.83**

The performance gap between MNIST and CIFAR-10 can be explained by the joint effect of the communication model and the learning task. According to the latency model in (23), the downlink dissemination time is determined by the global-model payload size and the achievable downlink rates of all devices, while the proposed RIS-assisted beamforming improves

the rates through MI maximization. Hence, for both datasets, a larger MI leads to a shorter dissemination time and supports more efficient FL training under the same wall-clock budget. Nevertheless, the benefit of faster dissemination is translated into learning accuracy through the local training dynamics. MNIST has simpler grayscale samples and lower feature diversity, so each successfully delivered global model can lead to relatively stable local updates and faster convergence. CIFAR-10, however, contains more complex RGB images with higher intra-class diversity and background variation. Therefore, even with the same communication improvement, the local updates on CIFAR-10 are more difficult to optimize and require more effective rounds or a more expressive model to reach the same performance level. As a result, the proposed method achieves consistent communication-learning gains on both datasets, but the absolute accuracy and F1-score on CIFAR-10 remain lower than those on MNIST.

VI. Conclusion

In this paper, we studied MI-driven communication design for a multi-user RIS-assisted FL system. By explicitly relating the downlink global model dissemination latency to the achievable MI, we formulated a joint optimization problem for the active beamforming at the BS and the passive phase shifts at the RIS, with the objective of maximizing the system sum MI under practical power and unit-modulus constraints. To handle the resulting non-convex coupling between the BS beamforming vectors and RIS phases, we developed an efficient AO framework that iteratively updates beamforming variables using tractable subproblem solvers and yields a monotonic improvement of the objective. Simulation results indicate that the proposed RIS-aided design improves communication-side metrics and is empirically associated with lower dissemination latency, improved communication–computation efficiency, and faster training progress under the considered setup. Overall, the proposed joint beamforming strategy provides a practical communication-side approach for improving the efficiency of RIS-assisted wireless FL.

References

[1] R. V. N. D. Silva, A. Flávia Dos Reis, G. Brante, R. D. Souza, "Hierarchical federated learning with distributed clustering and multichannel ALOHA," *IEEE Sens. J.*, vol. 25, no. 7, pp. 12128-12142, Apr. 2025.

[2] Q. Qi, X. Chen, "Robust design of federated learning for edge-intelligent networks," *IEEE Trans. Commun.*, vol. 70, no. 7, pp. 4469-4481, July 2022.

[3] Y. Sun, S. Zhou, Z. Niu, D. Gündüz, "Dynamic scheduling for over-the-air federated edge learning with energy constraints," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 1, pp. 227-242, Jan. 2022.

[4] M. Kim, A. L. Swindlehurst, D. Park, "Beamforming vector design and device selection in over-the-air federated learning," *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 7464-7477, Nov. 2023.

[5] S. H. Park, H. Lee, "Completion time minimization of fog-RAN-assisted federated learning with rate-splitting transmission," *IEEE Trans. Veh. Technol.*, vol. 71, no. 9, pp. 10209-10214, Sept. 2022.

[6] D. Xu, "Latency minimization for TDMA-based wireless federated learning networks," *IEEE Trans. Veh. Technol.*, vol. 73, no. 9, pp. 13974-13979, Sept. 2024.

[7] X. Deng, J. Li, C. Ma, *et al.*, "Low-latency federated learning with DNN partition in distributed industrial IoT networks," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 3, pp. 755-775, Mar. 2023.

[8] Z. Alsulaimawi, "Federated learning with privacy-preserving active learning: a min-max mutual information approach," *2024 Int. Conf. on Artif. Intell. in Inf. and Commun. (ICAIIC)*, Osaka, Japan, pp. 503-508, Feb. 2024.

[9] J. Zhang, X. Hu, Z. Cai, L. Zhu, Y. Feng, "Federated learning based on mutual information clustering for wireless traffic prediction," *Electron.*, vol. 12, no. 21, pp. 4476-4491, Oct. 2023.

[10] N. Chen, Y. Li, X. Liu, Z. Zhang, "A mutual information based federated learning framework for edge computing networks," *Comput. Commun.*, vol. 176, pp. 23-30, May 2021.

[11] M. Deng, M. Ahmed, A. Wahid, *et al.*, "Reconfigurable intelligent surfaces enabled vehicular communications: a comprehensive survey of recent advances and future challenges," *IEEE Trans. Intell. Veh.*, vol. 10, no. 8, pp. 4191-4216, Aug. 2025.

[12] L. Zhi, H. Niu, Y. He, *et al.*, "Self-powered absorptive reconfigurable intelligent surfaces for securing satellite-terrestrial integrated networks," *China Commun.*, vol. 21, no. 9, pp. 276-291, Sept. 2024.

[13] A. Chauhan, S. Ghosh, A. Jaiswal, "RIS partition-assisted non orthogonal multiple access (NOMA) and quadrature-NOMA with imperfect SIC," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4371-4386, July 2023.

[14] W. Zhu, Y. Zhang, W. Gao, Y. Xie, Y. Zhang, Z. Zhang, "Energy-efficient RIS-aided cell-free massive MIMO ISAC systems: joint AP operation mode selection and beamforming design," *IEEE Sens. J.*, vol. 25, no. 23, pp. 43402-43413, Dec. 2025.

[15] J. Ma, Q. Li, R. Liu, A. Pandharipande, X. Ge, "Enhanced semantic information transfer on RIS-assisted communication systems," *IEEE Wireless Commun. Lett.*, vol. 13, no. 8, pp. 2225-2229, Aug. 2024.

[16] H. Zhang, "Joint waveform and phase shift design for RIS-assisted integrated sensing and communication based on mutual information," *IEEE Commun. Lett.*, vol. 26, no. 10, pp. 2317-2321, Oct. 2022.

[17] H. D. Tuan, A. A. Nasir, E. Dutkiewicz, H. V. Poor, L. Hanzo, "RIS-aided multiple-input multiple-output broadcast channel capacity," *IEEE Trans. Commun.*, vol. 72, no. 1, pp. 117-132, Jan. 2024.

[18] H. Guo, Y. C. Liang, J. Chen, E. G. Larsson, "Weighted sum-rate maximization for reconfigurable intelligent surface aided wireless networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3064-3076, May 2020.

[19] Y. Han, S. Zhang, L. Duan, R. Zhang, "Double-IRS aided MIMO communication under LoS channels: capacity maximization and scaling," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2820-2837, Apr. 2022.

[20] M. H. Lee, H. Y. Lee, S. Y. Shin, "Reconfigurable intelligent surface-assisted time reversal multiple access," *IEEE Trans. Veh. Technol.*, vol. 74, no. 9, pp. 14828-14832, Sept. 2025.

[21] Q. Zhang, H. Zhou, W. Zhang, Y. C. Liang, H. V. Poor, "Channel capacity of RIS-assisted symbiotic radios with imperfect knowledge of channels," *IEEE Trans. Cognit. Commun. Networking*, vol. 10, no. 3, pp. 938-952, June 2024.

[22] Z. Wang, J. Qiu, Y. Zhou, *et al.*, "Federated learning via intelligent reflecting surface," *IEEE Trans. Wireless Commun.*, vol. 21, no. 2, pp. 808-822, Feb. 2022.

[23] A. M. Elbir, S. Coleri, "Federated learning for channel estimation in conventional and RIS-assisted massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 21, no. 6, pp. 4255-4268, June 2022.

[24] H. Liu, X. Yuan, Y. -J. A. Zhang, "Reconfigurable intelligent surface enabled federated learning: a unified communication-learning design approach," *IEEE Trans. Wireless Commun.*, vol. 20, no. 11, pp. 7595-7609, Nov. 2021.

[25] M. Chen, H. V. Poor, W. Saad, S. Cui, "Convergence time optimization for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 4, pp. 2457-2471, Apr. 2021.

[26] Q. Wu, R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394-5409, Nov. 2019.

[27] K. Zhong, J. Hu, H. Li, R. Wang, D. An, et al., "RIS-aided beamforming fesign for MIMO systems via unified manifold optimization," *IEEE Trans. Veh. Technol.*, vol. 74, no. 1, pp. 674-685, Jan. 2025.

[28] H. B. McMahan, E. Moore, D. Ramage, S. Hampson, B. A. Arcas, "Communication-efficient learning of deep networks from decentralized data," *20th Int. Conf. Artif. Intell. Stat. (AISTATS)*, Ft. Lauderdale, FL, USA, pp. 1273-1282, Apr. 2017.

[29] Y. Cai, S. Li, J. Zhang, D. Zhang, "RIS-assisted federated learning algorithm based on device selection and weighted averaging," *IEEE 99th Veh. Technol. Conf. (VTC2024-Spring)*, Singapore, Singapore, pp. 1-5, June 2024.

[30] C. Zhang, M. Dong, B. Liang, A. Afana, Y. Ahmed, "Multi-model wireless federated learning with downlink beamforming," *IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Seoul, Republic of Korea, pp. 9146-9150, Apr. 2024.

Yujun Cai (Graduate Student Member, IEEE) received his B.S. degree from Communication University of China, Beijing, China, in 2021. He is working toward the Ph.D. degree in information and communication engineering at the School of Information and Communication Engineering, Communication University of China.

His research interests include federated learning, reconfigurable intelligent surface, beamforming, and channel estimation.

Shufeng Li (Member, IEEE) received the B.Sc. and M.Sc. degrees from Hebei University of Technology, Tianjin, China, in 2004 and 2007, and Ph.D. degree from Beihang University, Beijing, China, in 2011.

He is currently a professor with the School of Information and Engineering, Communication University of China, Beijing, China. His research interests mainly include channel estimation, massive MIMO communication, and non-orthogonal multiple access.

Jianbo Liu received the bachelor's degree from the Department of Radio Electronics, Tsinghua University, Beijing, China, in 1985, and the master's degree from the Communication University of China, Beijing, in 1988.

He is currently a professor with the School of Information and Communication Engineering, Communication University of China. His research interests include cable TV and broadband network technology.

Shaolin Liao (Senior Member, IEEE) received the B.S. degree in materials science and engineering from Tsinghua University, Beijing, China, in 2000, and the Ph.D. degree in electrical engineering from the University of Wisconsin–Madison, Madison, WI, USA, in 2008.

Since 2021, he has been a full professor with Sun Yat-sen University, Guangzhou, China. He has been an adjunct faculty member with the Illinois Institute of Technology, Chicago, IL, USA, since 2017, and a visiting scholar with Purdue University, West Lafayette, IN, USA, since 2024. His research spans the multidisciplinary areas of AI-powered electromagnetics, photonics, and data science.

Xinruo Zhang (Member, IEEE) received the B. Eng degree in electronics engineering and satellite communications engineering from Beihang University, Beijing, China, in 2010, the M.Sc. degrees in electronics engineering from the University of Surrey, Guildford, U.K., in 2012, and the Ph.D. degree in telecommunications research from King's College London, London, U.K., in 2018.

She is currently a lecturer with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K. Her research includes machine learning for wireless communications, radio resource allocation, and mobile-edge intelligence.