Energy-Efficient Federated Learning With Intelligent Reflecting Surface

Ticao Zhang and Shiwen Mao, *Fellow, IEEE*

***Abstract*—Federated learning is a new paradigm to support resource-intensive and privacy-aware learning applications. It enables the Internet-of-Things (IoT) devices to collaboratively train a global model to accomplish a machine learning task without sharing private data. In practice, the IoT devices powered by batteries finish the local training and interact with the central server via wireless links. However, the repeated interaction between IoT devices and the central server would consume considerable resources. Motivated by the emerging technology of intelligent reflecting surface (IRS), we propose to leverage the IRS to reconfigure the wireless propagation environment to maximize the utilization of the available resources. Specifically, we consider the critical energy efficiency issue in the reconfigurable wireless communication network. We formulate an energy consumption minimization problem in an IRS-assisted federated learning system subject to the completion training time constraint. An iterative resource allocation algorithm is proposed to jointly configure the parameters with proven fast convergence. Simulation results validate that the proposed algorithm converges fast and can achieve significant energy savings, especially when the number of reflecting elements is large and when the IRS is properly configured.**

***Index Terms*—Federated learning, intelligent reflecting surface, energy-efficiency, resource allocation, Internet-of-Things (IoT).**

I. INTRODUCTION

THE EMERGING intelligent applications such as face recognition, autonomous driving, unmanned aerial vehicle (UAV), and indoor localization have imposed great challenges for Internet of Things (IoT) devices due to the computation-intensive and latency-sensitive features. The devices are generating a vast amount of data via their local sensors, e.g., GPS, accelerometer, and camera. It is envisioned that future networks should be able to utilize the local data at the mobile edge to perform intelligent inference and machine learning tasks. However, the paradigm change from "connected things" to "connected intelligence" in the era of 6G brought about two main challenges [1]. First, the bandwidth is limited, aggregating the large volumes of data would cause network congestion. Second, data-privacy is becoming a critical issue in today's IoT and the Internet. As a result, it becomes more and more desirable to perform learning tasks at the end-IoT devices instead of sending raw data to the central cloud.

A new machine learning method, termed federated learning, has emerged as a promising solution for privacy-sensitive and low-latency solutions [2]–[4]. In federated learning, user data is stored locally. In each communication round, users perform local training based on their local data and then upload their trained model to the central server. After aggregating the local updates from all users, the central server distributes the new global model to the users. This process proceeds in an iterative way until convergence is reached. In this way, a global model, which is trained from the data stored on each device, can be obtained without data leakage or data being inferred from other users. This property makes federated learning one of the most promising technologies of future intelligent networks.

Nevertheless, so far, the potential of federated learning has not been fully exploited yet due to the stochastic nature of wireless channels. For example, cell edge users often suffer from communication links of poor quality or unfavorable wireless propagation conditions. Fortunately, the recent advances in reconfigurable wireless technology provide a new cost-effective means to enhance the performance of intelligent learning systems [5], [6]. To be specific, the intelligent reflecting surface (IRS) is composed of a large number of reflecting elements, whose amplitude and phase can be adjusted to create a favorable propagation environment [7]–[9]. The direct channel gain in combination with the reflection-aided beamforming gain can boost the local model uploading performance.

In this paper, we investigate energy efficient communication in federated learning with IRS. There are several challenges. First of all, the IoT devices for federated learning are powered by batteries, which need to support both local training and model upload. How to save the battery power of each device becomes a critical issue. Second, the global model training accuracy depends on the number of training iterations. The parameters need to be properly designed to meet the training accuracy requirement while also conserve energy. Third, with the involvement of IRS, the parameters become highly coupled. A joint design of the IRS parameters as well as the computing/communication parameters is of critical importance. The main contributions of this paper include:

1) We investigate an IRS-assisted federated learning system, where the IRS reconfigures the communication channel so that the IoT devices can upload their model with a reduced power. As a result, the total energy consumption can be effectively reduced.

Manuscript received May 31, 2021; revised August 25, 2021; accepted November 2, 2021. Date of publication November 10, 2021; date of current version May 20, 2022. This work was supported in part by NSF under Grant ECCS-1923717 and Grant CNS-2107190, and in part by the Wireless Engineering Research and Education Center, Auburn University, Auburn, AL, USA. *(Corresponding author: Shiwen Mao.)*

The authors are with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849 USA (e-mail: tzz0031@auburn.edu; smao@ieee.org).

Digital Object Identifier 10.1109/TGCN.2021.3126795

2) We formulate a joint local training and model uploading problem, which aims to minimize the energy consumption subject to the task completion time requirement. A low complexity iterative algorithm with proven fast convergence is proposed to optimize each variable iteratively. Most of the variables can be obtained numerically with the simple one dimensional search algorithm, which makes it useful in practical systems. We show that the main complexity of the algorithm comes from the optimization of IRS elements, which involves solving an semidefinite programming (SDP) problem.
3) The convergence of the proposed algorithm is proved theoretically and verified numerically. Extensive simulations are performed to demonstrate the benefits brought by the use of IRS. Our results suggest that with the use of IRS, the energy consumption in federated learning of a battery-powered IoT device network can be greatly reduced, especially, when the number of reflecting elements is large and the IRS is properly configured.

The remainder of this paper is organized as follows. Section II introduces the relevant work and Section III presents the system model and problem statement. We start the design of the algorithm from the simplest case where there is only one device in Section IV. Then the algorithm is extended to a multi-device federated learning scenario in Section V and a low complexity algorithm is proposed in Section VI. Numerical results are discussed in Section VII. Finally, Section VIII concludes this paper.

Notation: The notation used in this paper is summarized as follows. Bold lower/upper case letters denote vectors and matrices, respectively. $\mathcal{CN}(\mu, \sigma^2)$ denotes the circularly symmetric complex Gaussian distribution with mean μ and variance σ^2. For any scalar a, $|a|$ denotes its absolute value. For any vector $\mathbf{a}$, a_i is the i-th element. $\mathbf{A}^*$, $\mathbf{A}^T$ and $\mathbf{A}^H$ represent the conjugate, transpose, and conjugate transpose of matrix $\mathbf{A}$, respectively. $\mathrm{Diag}(\mathbf{A})$ stands for a vector whose elements are extracted from the diagnal of matrix $\mathbf{A}$. $\mathbf{A} \succeq \mathbf{0}$ means that $\mathbf{A}$ is a positive semidefinite (PSD) matrix. $\mathrm{Rank}(\mathbf{A})$ denotes the rank of matrix $\mathbf{A}$. $\arg(\cdot)$ returns the angle of a complex variable. Variables with star indicate optimal solutions.

II. Related Works

A. Intelligent Reflecting Surface

IRS is an enabling technology to reconfigure the radio signal propagation in wireless links [7]–[9]. It has been regarded as a promising enabler for smart wireless communication for B5G/6G wireless systems. By deploying a large number of passive reflecting elements, the signal propagation channel can be smartly coordinated to achieve a desired distribution.

Earlier works suggest that a controllable surface could be realized by changing the electric and/or magnetic polarizability property of the scatter [10]. Later, this research area has been explored in terms of theoretical IRS signal and channel modeling [8], practical IRS beamforming design [6], and prototype deployment [11]. The beamforming design includes both passive beamforming at the IRS and active beamforming at the transmitter, which is optimized based on different objectives, such as power minimization [6], rate maximization [12], energy efficiency maximization [13], etc. Recently, IRS has been investigated for physical layer security [14], simultaneous power and energy transfer (SWIFT) [15], mobile edge computing [16], etc.

B. Energy Efficient Federated Learning

Federated learning, first proposed in [2], is a distributed learning method that enables IoT devices to train a global model without sharing their own data with other users. Due to its advantages in protecting privacy, it has been successfully adopted in a wide range of application scenarios, such as semantic location, health prediction, or learning sentiment [4].

There are a number of works focused on federated learning over wireless links. A communication and computation co-design approach for fast model aggregation is proposed in [17], which leverages the property of signal superimposition on wireless multiple access channels. This *over-the-air computation* (AirComp) framework is achieved by jointly considering the beamforming design and the device selection problem. A collaborative learning that takes into account of limited wireless resources is first investigated in [18]. The impact of MAC layer bandwidth and power limit on the performance of federated learning is investigated under the framework of AirComp. A general model that investigates the computation and communication latency trade-off in federated learning is proposed in [19]. The authors show that federated learning over wireless networks captures a trade-off between communication and computation. The previous research are all focused on stochastic wireless channels. The benefits of configurable technology such as IRS on the performance of federated learning has not been fully investigated. Recent results in mobile edge computing show that the overall uplink transmission latency can be reduced [20] and the system throughput can be improved [16] with the IRS technology.

There are several works that investigate federated learning with IRS. In [17], the authors show that when federated learning meets IRS, the model aggregation error can be reduced via the enhanced signal provided by the IRS. AirComp and IRS have the potential to tackle the challenge of the communication bottleneck problem. The authors in [21] investigate the model aggregation performance in a federated learning system with IRS. A joint model device selection, beamforming, and IRS phase shift optimization algorithm is proposed. The proposed algorithm can schedule more devices in each communication round under certain accuracy requirement.

III. System Model and Problem Formulation

As shown in Fig. 1, we consider a single-cell federated learning communication system, where K single antenna IoT devices offload their locally trained models to an edge server hosted at a BS with M antennas through radio access links. The federated learning model is the same as that in [22], where a global ML problem is solved at a central server with the training dataset partitioned over IoT devices.

We assume that each device k has a local training dataset with D_k data samples. The federated learning model is locally

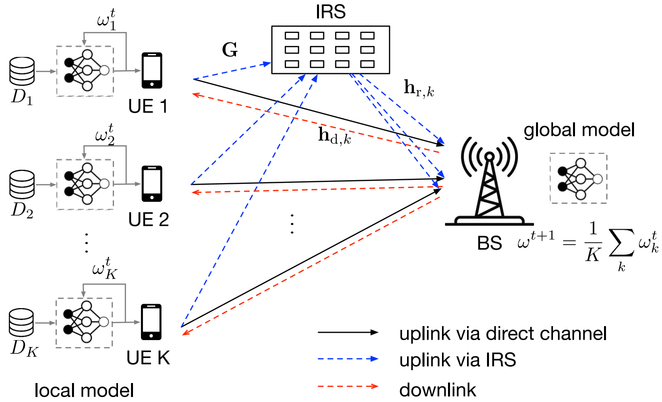

Fig. 1. Illustration of the federated learning system with IRS.

trained by each device's own dataset. Then the local model parameter is uploaded to the BS. After aggregation, the BS then broadcasts the global model to each participating device. This is called one round of training. Such communication round will be performed several times until the model achieves a required level of accuracy. *We aim to determine the resource allocation strategy to achieve an energy efficient design.*

A. Wireless Communication Model

We consider uplink frequency-division multiple access (FDMA) transmissions where the BS serves the users with orthogonal frequency bands. To assist the model uploading of mobile devices, an IRS with N reflecting elements is placed between the IoT devices and the BS. The equivalent channels from device k to the BS, from device k to the IRS, and from the IRS to the BS are denoted as $\mathbf{h}_{\mathrm{d},k} \in \mathbb{C}^{M\times 1}$, $\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{N\times 1}$ and $\mathbf{G} \in \mathbb{C}^{M\times N}$, respectively. The IRS has a reflection phase-shift matrix $\mathbf{\Theta} \in \mathbb{C}^{N\times N}$, which is a diagonal matrix with $e^{j\theta_n}$ being its diagonal elements, $\theta_n \in [0, 2\pi]$, for all $1 \leq n \leq N$. $\mathbf{\Theta}$ captures the effective phase shifts of all the reflecting elements of the IRS. The phase shift unit can be adjusted by the IRS controller based on measured channel dynamics. The composite channel is therefore modeled as a combination of the direct channel and the reflected channel. The training update transmission between the IoT device and the cloud server happens in orthogonal frequency bands. Hence there is no interference between users. Then the uplink transmission rate of the kth IoT device is given by

$$R_k = b_k \log_2\left(1 + \frac{p_k\left|\mathbf{w}_k^H \mathbf{h}_k\right|^2}{N_0\left|\mathbf{w}_k^H\right|^2 b_k}\right), \tag{1}$$

where b_k is the bandwidth allocated to device k, $\mathbf{n}$ is the additive white Gaussian noise (AWGN) with zero mean and noise power spectrum density N_0, $\mathbf{w}_k \in \mathbb{C}^{M\times 1}$ is the beamforming vector for device k, $\mathbf{h}_k \triangleq \mathbf{h}_{\mathrm{d},k} + \mathbf{G}\mathbf{\Theta}\mathbf{h}_{\mathrm{r},k} \in \mathbb{C}^{M\times 1}$ is the combined channel between device k and the BS.

B. Federated Learning Model

A federated learning process consists of three stages: local training, model aggregation, and model distribution. The entire training process differs from the conventional mobile edge computing system in three aspects. First of all, in mobile edge computing systems, a device can offload part of its work to the cloud while computing its own tasks asynchronously. However, for federated learning, each device has to finish its local model training first, and then performs model uploading. Second, in federated learning, the cloud cannot aggregate the global model until each device offloads its local model to the cloud. This requires stringent synchronous processing and poses the latency requirement. This training process usually lasts for several rounds. Third, in federated learning, the uploaded model sizes should be the same for all the IoT devices, while the uploaded data sizes are usually different across different devices in general mobile edge computing. The models for the three stages of federate learning are provided in the following.

1) Local Training: When an application is executed on the IoT device, the energy consumption depends on the CPU workload of the device, which is characterized by the number of CPU cycles to complete this application. Assume c_k is the number of CPU cycles required to process one bit and f_k is the number of CPU cycles per second for device k. Then the time required for carrying out the local model training can be expressed as $D_k c_k/f_k$ in each local training round. We assume that each device uses the stochastic average gradient (SAG) algorithm to train the local model to achieve a local level relative accuracy $\eta \in [0, 1]$. The number of local iterations is then given by [23]

$$L(\eta) = \ell_1 \ln(1/\eta), \tag{2}$$

where $\ell_1 > 0$ is a parameter depending on the data size and structure of the local problem. In [24], it is shown that the local level accuracy $\eta = 0$ describes an exact solution of the subproblem and $\eta = 1$ means that the local training has not been improved at all. In this case, the local training latency will be

$$t_k^L = L(\eta) D_k c_k / f_k. \tag{3}$$

Assume the IoT device uses a dynamic voltage scaling (DVS) scheme, so it can adjust its computational speed to save energy [25]. According to [25], the energy consumption per CPU cycle can be expressed as κf_k^2, where κ is a coefficient depending on the chip architecture. Then the energy consumption for local training can be expressed as

$$E_k^L = \kappa D_k c_k f_k^2 L(\eta). \tag{4}$$

2) Model Aggregation: After local model training, each IoT device then sends its local updates to the BS. Suppose S is the size of the offloading training model with a fixed dimension, which should be the same for all the IoT devices. The upload latency can be expressed as

$$t_k^U = S/R_k, \tag{5}$$

where R_k is given in (1). The energy consumption of model uploading for device k is expressed as

$$E_k^U = t_k^U\left(p_{c,k} + p_k\right), \tag{6}$$

where $p_{c,k}$ is a constant circuit power of the IoT device during the computational uploading process.

3) Model Distribution: The parameters related to the global model are updated via a simple linear processing at the cloud server hosted at the BS. The BS has a strong processing capability, and hence the processing time can be negligible. After the global model parameters are updated at the BS, the BS distributes the global model parameters to all the IoT devices. The broadcast time can also be negligible since the BS has high transmit power and large bandwidth.

To achieve a global accuracy ϵ, the number of global iterations is given by [24]

$$G(\eta) = \frac{\mathcal{O}\big(\ln\big(\frac{1}{\epsilon}\big)\big)}{1-\eta}. \tag{7}$$

In this work, we consider a fixed, target global accuracy ϵ, so we can normalize $\mathcal{O}(\ln(\frac{1}{\epsilon}))$ to 1 without changing the nature of this problem.

To this end, the overall latency of IoT device k is composed of the local computation time and model uploading latency as

$$T_k = \frac{1}{1-\eta}\Big(t_k^L + t_k^U\Big). \tag{8}$$

Let T be the maximum training time for the entire federated learning algorithm. Then we have

$$T_k \le T,\ \forall k. \tag{9}$$

The overall energy consumption for IoT device k over the entire federated learning process is

$$E_k = \frac{1}{1-\eta}\Big(E_k^L + E_k^U\Big). \tag{10}$$

C. Problem Formulation

To allow the IoT devices to save energy while also guaranteeing the training time/accuracy requirements of federated learning, we need to develop effective resource allocation algorithms. The energy minimization problem is thus formulated as follows.

$$\begin{aligned} (P1) \quad & \min_{\eta, f_k, b_k, p_k, \mathbf{w}_k, \boldsymbol{\Theta}} \sum_k E_k \\ & \text{s.t.} \quad C1\colon T_k \le T,\ \forall k \\ & \qquad\ \ C2\colon \eta \ge 0 \\ & \qquad\ \ C3\colon 0 \le \theta_n \le 2\pi,\ \forall n \\ & \qquad\ \ C4\colon 0 \le p_k \le P_{\max},\ \forall k \\ & \qquad\ \ C5\colon \sum_k b_k \le B, \end{aligned} \tag{11}$$

where constraint (C1) is the task completion time constraint; (C2), (C3), and (C4) specify the domain of η, θ_n, and p_k, respectively; constraint (C5) indicates that the combined occupied bandwidth should not exceed the total available bandwidth. This is a joint power, bandwidth, phase shift, accuracy control, and beamforming design problem. Problem (P1) has a non-convex and mixed structure where some variables are coupled. Obtaining a global optimal solution will be quite challenging.

IV. Analysis of the Single Device System

First of all, we consider the simplest case where there is only one IoT device. Although such assumption is not practical in terms of federated learning, the results can still provide useful insights on parameter optimization for a practical multiuser federated learning system. In the rest of this section, we set $k = 1$. The total energy consumption for device k is

$$E_k = \frac{1}{1-\eta}\bigg(\frac{S}{R_k}\big(p_{c,k} + p_k\big) + \kappa D_k c_k f_k^2 L(\eta)\bigg). \tag{12}$$

A. Design of the Device CPU Frequency

Theorem 1: The optimal operating frequency for device k is given by

$$f_k^* = \frac{L(\eta) D_k c_k}{T/G(\eta) - S/R_k}. \tag{13}$$

Proof: The objective function E_k in (12) is an increasing function in terms of f_k. The time constraint (C1) of Problem (P1) suggests that the IoT device should work on the lowest frequency f_k^* that is allowed by the delay constraint. ■

B. Design of Power Allocation

Next, we substitute the optimal solution f_k^* (13) into the original Problem (P1). We jointly optimize the power allocation when the local accuracy parameter η, the bandwidth b_k, the IRS parameters $\boldsymbol{\Theta}$ and $\mathbf{w}_k$ are known. The objective function becomes

$$E_k = G(\eta)\Bigg(\frac{S}{R_k}\big(p_{c,k} + p_k\big) + \kappa D_k c_k L(\eta)\bigg(\frac{L(\eta) D_k c_k}{T/G(\eta) - S/R_k}\bigg)^2\Bigg), \tag{14}$$

where R_k is a function of p_k, b_k, and $\boldsymbol{\Theta}$. A direct optimization is quite hard. To solve this problem, we optimize each variable in an iterative manner. Specifically, we write

$$f_k^2 = \frac{f_k^{t,3}}{f_k^*} = \frac{f_k^{t,3}}{L(\eta) D_k c_k}\bigg(\frac{T}{G(\eta)} - \frac{S}{R_k}\bigg), \tag{15}$$

where f_k^t is the result in the tth iteration. The objective function (12) then assumes a simpler form as $E_k = G(\eta)(\frac{S}{R_k}(p_{c,k} + p_k) + \kappa f_k^{t,3}(\frac{T}{G(\eta)} - \frac{S}{R_k}))$. When f_k is fixed, the problem becomes

$$\begin{aligned} (P2a) \quad & \min_{p_k} \frac{p_k + p_{c,k} - A_k}{R_k} \\ & \text{s.t.} \ \ (C4), \end{aligned} \tag{16}$$

where $A_k = \kappa f_k^{t,3}$ is a constant in each iteration step.

Theorem 2: The optimal solution to (P2a) when $p_{c,k} - A_k > 0$ is given by

$$p_k^* = \min\{p_k', P_{\max}\}, \tag{17}$$

where p_k' is the solution to $h(p_k) = \frac{a_k}{b_k + a_k p_k}(p_{c,k} + p_k - A_k) - \ln(1 + a_k p_k/b_k) = 0$.

Proof: If $p_{c,k} - A_k > 0$, then $p_k + p_{c,k} - A_k > 0$. Minimizing the energy consumption E_k is equivalent to maximizing the function $g(p_k) = \frac{R_k}{p_{c,k} + p_k - A_k}$. For simplicity of notation, we rewrite $g(p_k)$ as

$$g(p_k) = \frac{b_k}{\ln(2)} \frac{\ln(1 + a_k p_k/b_k)}{p_{c,k} + p_k - A_k}, \tag{18}$$

where $a_k = \frac{|\mathbf{w}_k^H \mathbf{h}_k|^2}{N_0 |\mathbf{w}_k^H|^2} > 0$. Then we have

$$g'(p_k) = \frac{b^k}{\ln(2)} \frac{\frac{a_k}{b_k + a_k p_k}\left(p_{c,k} + p_k - A_k\right) - \ln(1 + a_k p_k / b_k)}{\left(p_{c,k} + p_k - A_k\right)^2}. \tag{19}$$

Let the numerator be denote by $h(p_k)$ and we have

$$h'(p_k) = \frac{-a_k^2}{\left(b_k + a_k p_k\right)^2}\left(p_{c,k} + p_k - A_k\right) < 0, \tag{20}$$

which means $h(p_k)$ is a decreasing function on $[0, P_{\max}]$. Also note that $h(0) = a_k(p_{c,k} - A_k)/b_k > 0$ and $\lim_{p_k \to \infty} h(p_k) = -\infty$. Hence there exists a $p_k' \in [0, \infty]$ such that $h(p_k') = 0$. As a result, $h(p_k) > 0$ on the interval $[0, p_k']$ and $h(p_k) < 0$ on the interval $[p_k', +\infty]$. Hence $g'(p_k) > 0$ on the interval $[0, p_k']$ and $g'(p_k) < 0$ on the interval $[p_k', +\infty]$. We then claim that $g(p_k)$ achieves its maximum value when $p_k = p_k'$.

It is not straightforward to obtain a closed-form expression of p_k^* by solving $h(p_k) = 0$. However, this is a one-dimensional search problem and function $h(p_k)$ has the monotone property. Hence, some simple algorithms (e.g., bisection search) can be used to obtain the solution [26]. ∎

If $p_{c,k} - A_k \leq 0$, then the objective function in the subproblem is negative when $p_k + p_{c,k} - A_k < 0$ and positive when $p_k + p_{c,k} - A_k > 0$. By investigating the monotonicity of the objective function, we find that $\frac{R_k}{A_k - p_{c,k} - p_k}$ is strictly increasing on the interval $[p_{k,\min}, P_{\max}]$. Hence the objective function is minimized when $p_k = p_{k,\min}$. In this paper, we only consider the case where $p_{c,k} - A_k > 0$ for simplicity. The case $p_{c,k} - A_k < 0$ can be similarly analyzed.

C. Design of Bandwidth Allocation and IRS Parameters

When the power and frequency parameters are fixed, we can see that minimizing energy consumption is equivalent to maximizing the achievable rate R_k. The subproblem becomes

$$(P2b) \quad \max_{b_k, \mathbf{w}_k, \mathbf{\Theta}} \; R_k \quad \text{s.t.} \; (C3), (C5). \tag{21}$$

First of all, we can prove that R_k is a concave function w.r.t. b_k on the interval [0, *B*]. The optimal bandwidth allocation b_k for the singe device case can also be obtained with a bisection method by setting the first derivative of R_k to zero. To save space, we leave out this part of content.

Now, we optimize the IRS related parameters $\mathbf{\Theta}$ and $\mathbf{w}_k$. Maximizing the achievable rate R_k is equivalent to maximizing the corresponding SNR a_k. The problem becomes

$$\max_{\theta_n, \mathbf{w}_k} \; \frac{p_k \left|\mathbf{w}_k^H \mathbf{h}_k\right|^2}{N_0 \left|\mathbf{w}_k^H\right|^2 b_k} \tag{22a}$$

$$\text{s.t.} \; 0 \leq \theta_n \leq 2\pi, \tag{22b}$$

where $\mathbf{h}_k \triangleq \mathbf{h}_{\mathrm{d},k} + \mathbf{G\Theta h}_{\mathrm{r},k}$.

This problem can be solved by alternative optimization. Specifically, we first fix the IRS phase shift matrix $\mathbf{\Theta}$ and find the optimal detection vector $\mathbf{w}_k$. Without changing the nature of the problem, one can set $|\mathbf{w}_k^H|^2 = 1$ for simplicity. This problem becomes the well-known maximum ratio combing (MRC) detection problem. The SNR is maximized at

$$\mathbf{w}_k^* = \frac{\mathbf{h}_k}{|\mathbf{h}_k|}. \tag{23}$$

Next, for a fixed $\mathbf{w}_k^H$, we optimize the IRS phase vector. This problem is equivalent to the following problem.

$$\max \; \left|\mathbf{w}_k^H\left(\mathbf{h}_{\mathrm{d},k} + \mathbf{G\Theta h}_{\mathrm{r},k}\right)\right| \tag{24a}$$

$$\text{s.t.} \; 0 \leq \theta_n \leq 2\pi, \tag{24b}$$

We follow a similar procedure as in [6] by rewriting

$$\begin{aligned}\left|\mathbf{w}_k^H\left(\mathbf{h}_{\mathrm{d},k} + \mathbf{G\Theta h}_{\mathrm{r},k}\right)\right| &\leq \left|\mathbf{w}_k^H \mathbf{h}_{\mathrm{d},k}\right| + \left|\mathbf{w}_k^H \mathbf{G\Theta h}_{\mathrm{r},k}\right| \\ &\leq \left|\mathbf{w}_k^H \mathbf{h}_{\mathrm{d},k}\right| + \left|\mathbf{w}_k^H \mathbf{G}\mathrm{diag}\left(\mathbf{h}_{\mathrm{r},k}\right)\right|,\end{aligned} \tag{25}$$

where the first inequality is due to the triangle inequality and the equality holds if and only if $\arg(\mathbf{w}_k^H \mathbf{h}_{\mathrm{d},k}) = \arg(\mathbf{w}_k^H \mathbf{G\Theta h}_{\mathrm{r},k}) \triangleq \phi_0$. Note that $\mathrm{diag}(\mathbf{\Theta})$ is a diagonal matrix and we extract its diagonal as a vector $\mathbf{v} = \mathrm{diag}(\mathbf{\Theta})$, then $\mathbf{w}_k^H \mathbf{G\Theta h}_{\mathrm{r},k} = \mathbf{w}_k^H \mathbf{G}\mathrm{diag}(\mathbf{h}_{\mathrm{r},k})\mathbf{v}$. Considering the constraint that $|v_n| = |e^{j\theta_n}| = 1$, the optimal solution to this problem is given by

$$\mathbf{v}^* = \exp\left\{j\left(\phi_0 - \arg\left(\mathbf{w}_k^H \mathbf{G}\mathrm{diag}\left(\mathbf{h}_{\mathrm{r},k}\right)\right)\right)\right\}, \tag{26}$$

where $\phi_0 = \arg(\mathbf{w}_k^H \mathbf{h}_{\mathrm{d},k})$.

Remark 1: It can be seen that the IRS can strengthen the received signal power by aligning the cascaded channel with the direct channel compared with that without IRS.

This alternating optimization method is appealing since it has a closed-form expression for both the IRS phase shift vector and the signal detection vector. Its convergence is guaranteed since each subproblem ensures that the objective function is non-decreasing over iterations and is bounded above as the second inequality in (25) suggests.

D. Design of the Accuracy Parameter

Finally, we optimize the accuracy parameter η. For simplicity of notation, the objective function can be rewritten as

$$f(\eta) = \frac{1}{1-\eta}\left(u + v\log\left(\frac{1}{\eta}\right)\right), \; \eta \in (0, 1), \tag{27}$$

where $u = (p_{c,k} + p_k)S/R_k$ and $v = \kappa D_k c_k f_k^2 \ell_1$ are both positive numbers.

Theorem 3: The optimal accuracy parameter η^* is the solution to $h(\eta) = 0$, where

$$h(\eta) = -v(1-\eta) + u\eta - v\eta\ln(\eta). \tag{28}$$

Proof: Let's examine the property of $f(\eta)$. Its first order derivative is

$$f'(\eta) = \frac{-v(1-\eta) + u\eta - v\eta\ln(\eta)}{\eta(1-\eta)^2}. \tag{29}$$

Algorithm 1 Energy-Efficient Optimization for Single Device

1: Initialize IRS phase shift matrix $\boldsymbol{\Theta}$ and the iteration number $t = 1$, $s = 1$;
2: **repeat**
3: Obtain $\mathbf{w}_k^s$ according to (23);
4: Obtain $\mathbf{v}^s$ according to (26);
5: $s = s+1$;
6: **until** convergence
7: **repeat**
8: Obtain f_k^t according to (13);
9: Obtain p_k^t according to (17);
10: Obtain b_k^t ;
11: Obtain η^t according Theorem 3;
12: Calculate E_k based on (12);
13: $t = t+1$;
14: **until** $\frac{|E_k^{t+1} - E_k^t|}{|E_k^t|} \leq \epsilon_1$ and (C1) is satisfied

Letting $h(\eta) = -v(1-\eta) + u\eta - v\eta \ln(\eta)$, we have $\lim_{\eta\to 0^+} h(\eta) = -v < 0$ and $\lim_{\eta \to 1^-} h(\eta) = u > 0$. Moreover,

$$h'(\eta) = u - v\ln(\eta) > 0, \ \eta \in (0, 1). \tag{30}$$

Hence, $h(\eta)$ is an increasing function in terms of η on (0, 1) and there exists only one point η' such that $h(\eta') = 0$. Then $h(\eta) < 0$ ($f'(\eta) < 0$) on the interval $(0, \eta']$, and $h(\eta) > 0$ ($f'(\eta) > 0$) on the interval $[\eta', 1)$. As a result, $f(\eta)$ will be decreasing on $(0, \eta']$ and increasing on $[\eta', 1)$. The optimal solution that minimizes the energy consumption E_k should be $\eta = \eta'$. ∎

The algorithm for single user training is presented in Algorithm 1. The mainly complexity comes from the alternative updates of $\boldsymbol{\Theta}$ and $\mathbf{w}_k$, whose complexity are on the order of $\mathcal{O}(MN, N^2)$ and of $\mathcal{O}(MN^2)$, respectively. We can therefore claim that the overall complexity is $\mathcal{O}(I_1 MN^2)$, where I_1 is the iteration involved in Lines 2-6 in Algorithm 1.

V. Analysis of the Multiuser Federated Learning System

In this section, we consider the more practical multiuser federated learning system. The objective function becomes

$$\begin{aligned} E &= \sum_k E_k \\ &= \sum_k G(\eta)\bigg(\frac{S}{R_k}\big(p_{c,k} + p_k\big) + \kappa D_k c_k f_k^2 L(\eta)\bigg). \end{aligned} \tag{31}$$

A. Design of the Device CPU Frequency

First of all, we optimize the frequency when the training accuracy η is known. Minimizing the sum energy consumption of each device is equivalent to minimizing the individual energy consumption of each device. Again, for each device, E_k is an increasing function in terms of f_k. As a result, the frequency should be set as (13) to satisfy the latency constraint of each device.

B. Design of Power Allocation

From the objective function (31), we find that minimizing the energy consumption for all users is equivalent to

$$\begin{aligned} (P3a) \quad &\min_{p_k} \ \sum_k \frac{p_k + p_{c,k} - A_k}{R_k} \\ &\text{s.t.} \ \ (C3), \end{aligned} \tag{32}$$

where $R_k = b_k \log_2(1 + \frac{p_k |\mathbf{w}_k^H \mathbf{h}_k|^2}{N_0 |\mathbf{w}_k^H|^2 b_k})$. Similarly, minimizing the sum of energy consumption is equivalent to minimizing the energy consumption of each device. Hence the optimal power allocation can be similarly obtained as (17), which is a one dimensional search problem for each user.

C. Joint Design of Bandwidth Allocation and IRS Parameters

When the power and frequency are fixed in the last iteration, it is easy to verify that the optimal detection vector should be the same as the single device case as in (23), which maximizes the SNR for each device. Hence, the problem becomes

$$\begin{aligned} (P3b) \quad &\min_{\boldsymbol{\Theta}, b_k} \ \sum_k \frac{p_{c,k} + p_k - A_k}{b_k \log_2\big(1 + p_k |\mathbf{h}_k^H \mathbf{h}_k|^2 / (N_0 b_k)\big)} \\ &\text{s.t.} \ \ (C3), (C5), \end{aligned} \tag{33}$$

where $\mathbf{h}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{G}\boldsymbol{\Theta}\mathbf{h}_{\mathrm{r},k}$.

The problem is difficult since the variables b_k and $\boldsymbol{\Theta}$ are coupled in the numerator and the problem is non-convex. Moreover, the objective function in (P3b) is still not straightforward with the phase shift vector $\boldsymbol{\Theta}$. Now we extract the diagonal elements of $\boldsymbol{\Theta}$ to have $\bar{\mathbf{v}} = \mathrm{diag}\{\boldsymbol{\Theta}\} \in \mathbb{C}^{N\times 1}$. Supposing $\mathbf{H}_k = \mathbf{G}\mathrm{diag}\{\mathbf{h}_{\mathrm{r},k}\} \in \mathbb{C}^{M\times N}$, we have

$$\mathbf{h}_k = \mathbf{h}_{\mathrm{d},k} + \mathbf{H}_k \bar{\mathbf{v}} \tag{34}$$

$$\begin{aligned} \left|\mathbf{h}_k^H \mathbf{h}_k\right|^2 &= \mathbf{h}_{\mathrm{d},k}^H \mathbf{h}_{\mathrm{d},k} + \bar{\mathbf{v}}^H \mathbf{H}_k^H \mathbf{H}_k \bar{\mathbf{v}} \\ &\quad + \mathbf{h}_{d,k}^H \mathbf{H}_k \bar{\mathbf{v}} + \bar{\mathbf{v}}^H \mathbf{H}_k^H \mathbf{h}_{d,k}. \end{aligned} \tag{35}$$

By introducing an auxiliary matrix $\mathbf{R}_k \in \mathbb{C}^{(N+1)\times(N+1)}$ and an auxiliary vector $\mathbf{v} \in \mathbb{C}^{(N+1)\times 1}$, we further obtain

$$\mathbf{R}_k = \begin{bmatrix} \mathbf{H}_k^H \mathbf{H}_k & \mathbf{H}_k^H \mathbf{h}_{d,k} \\ \mathbf{h}_{d,k}^H \mathbf{H}_k & 0 \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \bar{\mathbf{v}} \\ 1 \end{bmatrix}. \tag{36}$$

Eqn. (35) can be further simplified as

$$\begin{aligned} \left|\mathbf{h}_k^H \mathbf{h}_k\right|^2 &= \mathbf{v}^H \mathbf{R}_k \mathbf{v} + \mathbf{h}_{\mathrm{d},k}^H \mathbf{h}_{\mathrm{d},k} = \mathrm{Tr}(\mathbf{R}_k \mathbf{V}) + \mathbf{h}_{\mathrm{d},k}^H \mathbf{h}_{\mathrm{d},k} \\ &\triangleq f_k(\mathbf{V}), \forall k \in \mathcal{K}, \end{aligned} \tag{37}$$

where $\mathbf{V} = \mathbf{v}\mathbf{v}^H \in \mathbb{C}^{(N+1)\times(N+1)}$. Then Problem (P3b) is equivalently transformed to (P3c), given by

$$(P3c) \quad \min_{\mathbf{V}, b_k} \ \sum_k \frac{p_{c,k} + p_k - A_k}{b_k \log_2(1 + p_k f_k(\mathbf{V})/(N_0 b_k))} \tag{38a}$$

$$\text{s.t.} \ \ V_{n,n} = 1 \tag{38b}$$

$$\mathrm{rank}(\mathbf{V}) = 1 \tag{38c}$$

$$\mathbf{V} \succeq \mathbf{0} \tag{38d}$$

$$(C5).$$

Note that constraints (38b) and (38d) ensure that $\mathbf{V} = \mathbf{v}\mathbf{v}^H$ holds true after optimization. Constraint (38c) is introduced to guarantee the unit modulus constraint when recovering $\mathbf{v}$ from $\mathbf{V}$. Due to the rank one constraint, this problem is non-convex in terms of the optimization variable $\mathbf{V}$. However, we have the following theorem.

Theorem 4: After dropping the rank one constraint (38c), Problem (P3c) is convex in terms of b_k and $\mathbf{V}$, respectively.

Proof: The objective function in Problem (P3c) is in the form of sum-of-ratios. To prove the sum function is convex, we only need to show that each sub ratio function is convex.

First of all, we consider the phase shift matrix $\mathbf{V}$. It can be seen that the denominator is actually a logarithm function of $f_k(\mathbf{V})$, which is concave, and $f_k(\mathbf{V})$ is a linear function of $\mathbf{V}$. Hence, each individual ratio function is convex in terms of $\mathbf{V}$. Hence, their sum will also be convex in terms of $\mathbf{V}$.

Next, note that the denominator of each sub ratio function is the achievable rate $R_k(b_k)$ of each device. Since

$$\frac{\partial^2 R_k}{\partial b_k^2} = -\frac{p_k^2 f_k(\mathbf{V})^2}{\ln(2)(N_0 b_k + p_k f_k(\mathbf{V}))^2 b_k} < 0, \tag{39}$$

$R_k(b_k)$ is concave in terms of b_k and $(p_{c,k} + p_k)/R_k(b_k)$ is convex in terms of b_k. ■

Note that simply dropping the rank one constraint (38c) does not necessarily result in an optimal $\mathbf{v}^*$ due to the additional constraint (38b). In other words, the optimal solution to Problem (P3c) after dropping the rank one constraint might not be feasible. One way is to use the Gaussian randomization method as shown in [6] to find an approximated solution. A common way to recover $\mathbf{v}$ from $\mathbf{V}$ is to denote $\mathbf{V}$ as $\mathbf{V} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^H$, where $\mathbf{U} \in \mathbb{C}^{(N+1)\times(N+1)}$ is an unitary matrix and $\boldsymbol{\Sigma}$ is an eigenvalue diagonal matrix. A feasible solution is constructed as $\hat{\mathbf{v}} = \mathbf{V}\boldsymbol{\Sigma}^{1/2}\boldsymbol{\zeta}$, where $\boldsymbol{\zeta} \in \mathbb{C}^{(N+1)\times 1}$ is a randomly generated complex circularly symmetric Gaussian random variable with zero mean and unit variance. The solution can be recovered by $\mathbf{v}^* = \exp\{j \arg(\frac{\mathbf{v}}{v_{N+1}})\}$ where v_{N+1} is the last element of vector $\mathbf{v}$. The optimal $\mathbf{V}^*$ can be further obtained from $\mathbf{v}^*$.

We denote the problem of (P3c) after dropping the rank one constraint (38c) as problem (P3c2'). Since this problem is in the form of sum-of-ratios, conventional fractional programming techniques such as the Dinkelbach's method cannot be used. To solve this problem, we first transform Problem (P3c2') into its equivalent form (P3d) by introducing auxiliary variable β_k.

$$(P3d) \quad \min_{\mathbf{V},\beta_k,b_k} \sum_k \beta_k \tag{40}$$

$$\text{s.t.} \quad \frac{p_{c,k} + p_k - A_k}{b_k \log_2(1 + p_k f_k(\mathbf{V})/(N_0 b_k))} \leq \beta_k$$
$$\text{(38b)}, \text{(38d)}, (C5). \tag{41}$$

Theorem 5: If $\mathbf{V}^*$, $\{b_k^*\}$, and $\{\beta_k^*\}$ are the optimal solution to (P3d), then there exists $\{\lambda_k^*\}$ such that $\mathbf{V}^*$ and $\{b_k^*\}$ are a solution to the following problem for $\lambda_k = \lambda_k^*$ and $\beta_k = \beta_k^*$.

$$(P3d') \min_{\mathbf{V},b_k} \sum_k \lambda_k \bigg(p_{c,k} + p_k - A_k - \beta_k b_k \log_2\bigg(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\bigg)\bigg)$$
$$\text{s.t.} \quad \text{(38b)}, \text{(38d)}, (C5), \tag{42}$$

and $\mathbf{V}^*$, $\{b_k^*\}$ also satisfy the following system equation for $\lambda_k = \lambda_k^*$ and $\beta_k = \beta_k^*$:

$$\lambda_k = \frac{1}{b_k^* \log_2\Big(1 + \frac{p_k f_k(\mathbf{V}^*)}{N_0 b_k^*}\Big)} \tag{43a}$$

$$\beta_k = \frac{p_{c,k} + p_k - A_k}{b_k^* \log_2\Big(1 + \frac{p_k f_k(\mathbf{V}^*)}{N_0 b_k^*}\Big)}. \tag{43b}$$

Proof: Theorem 4 also suggests that Problem (P3d') is a convex optimization problem. Introduce the Lagrange multipliers associated with the objective function in (P3d') and the bandwidth constraint. Thus the Lagrange function of the problem can be written as $\mathcal{L} = \sum_k \beta_k + \sum_k \lambda_k (p_{c,k} + p_k - A_k - \beta_k b_k \log_2(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k})) + \mu(\sum_k b_k - B)$. Then the optimal solution $\mathbf{V}^*$, $\{b_k^*\}$, and $\{\beta_k^*\}$ and the Lagrange multipliers λ_k^* and μ^* should satisfy the following KKT conditions.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{V}} = -\frac{1}{\ln 2} \sum_k \lambda_k \beta_k b_k \frac{1}{1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}} f_k'(\mathbf{V}) = 0 \tag{44a}$$

$$\frac{\partial \mathcal{L}}{\partial b_k} = \mu + \frac{\lambda_k \beta_k}{\ln(2)(N_0 b_k + p_k f_k(\mathbf{V}))} \bigg[p_k f_k(\mathbf{V}) - (p_k f_k(\mathbf{V}) + N_0 b_k) \ln\bigg(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\bigg)\bigg] = 0 \tag{44b}$$

$$\frac{\partial \mathcal{L}}{\partial \beta_k} = 1 - \lambda_k b_k \log_2\bigg(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\bigg) = 0 \tag{44c}$$

$$\lambda_k \frac{\partial \mathcal{L}}{\partial \lambda_k} = \lambda_k \bigg(p_{c,k} + p_k - A_k - \beta_k b_k \log_2\bigg(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\bigg)\bigg) = 0 \tag{44d}$$

$$\mu \frac{\partial \mathcal{L}}{\partial \mu} = \mu \bigg(\sum_k b_k - B\bigg) = 0 \tag{44e}$$

$$\lambda_k \geq 0, \mu \geq 0. \tag{44f}$$

$$p_{c,k} + p_k - A_k - \beta_k b_k \log_2\bigg(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\bigg) \leq 0 \tag{44g}$$

$$\sum_k b_k - B \leq 0. \tag{44h}$$

From (44c), we can infer that $\lambda_k^* > 0$ and conclude that the equality in (43a) holds. Similarly, from (44d), we have (43b). Moreover, note that given $\lambda = \lambda_k^*$ and $\beta_k = \beta_k^*$, (44a), (44b), (44e), and (44f) are just the KKT conditions for Problem (P3). Since (P3) is convex programming for parameter $\lambda_k > 0$ and $\beta_k \geq 0$, the KKT conditions are also sufficient optimality conditions. This completes the proof of the theorem. ■

Theorem 5 shows that the solution to Problem (P3d) can be obtained by finding the solutions that satisfy the KKT conditions in (44) among the solutions to Problem (P3d'). More important, if the solution is unique, it will be the global optimal solution.

1) Finding $(\mathbf{V}^*, b_k^*)$ When λ_k and β_k Are Given: Note that for a fixed λ_k and β_k, Problem (P3) belongs to convex optimization. Hence effective algorithms can be designed to find the optimal solution $(\mathbf{V}^*, b_k^*)$. Since $\mathbf{V}$ and b_k are coupled, we propose to use the alternative optimization which optimizes each variable alternatively. First of all, when b_k is known, Problem (P3d') becomes an SDP, which can be solved with existing optimization tools such as CVX[27].

$$\begin{aligned}\max_{\mathbf{V}} \quad & \sum_k \lambda_k \beta_k b_k \log_2\left(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\right) \\ \text{s.t.} \quad & (38\text{b}), (38\text{d}).\end{aligned} \tag{45}$$

Theorem 6: The optimal solution of b_k to (P3d') is given by

$$b_k = \frac{p_k f_k(\mathbf{V})}{N_0 x_k}, \tag{46}$$

where

$$x_k = -\frac{1}{W_L\left(-e^{-C_k}\right)} - 1, \tag{47}$$

and $W_L(\cdot)$ is the *Lambert W* function. Note that the μ in the expression of C_k is the Lagrange multiplier for Problem (P3d') satisfying $\sum_k b_k = B$.

Proof: Following Theorem (5), the optimal solution to (P3d') should satisfy the KKT conditions (44). We have from (44b)

$$\mu = \frac{\lambda_k \beta_k \left(-p_k f_k(\mathbf{V}) + (N_0 b_k + p_k f_k(\mathbf{V})) \ln\left(1 + \frac{p_k f_k(\mathbf{V})}{N_0 b_k}\right)\right)}{(N_0 b_k + p_k f_k(\mathbf{V})) \ln(2)},$$

which can be written as

$$\frac{\lambda_k \beta_k}{\ln(2)} \left(\ln(1+x) - \frac{x}{1+x}\right) = \mu, \tag{48}$$

where $x = \frac{p_k f_k(\mathbf{V})}{N_0 b_k}$. The solution is found to be

$$\ln(1+x) + \frac{1}{1+x} = 1 + \frac{\mu \ln(2)}{\lambda_k \beta_k} \triangleq C_k. \tag{49}$$

Hence, we obtain (46) and (47). Note that the optimal solution of bandwidth b_k^* can be obtained numerically by substituting μ^* into (47) and (46). Again, the bisection algorithm can be applied to find the numerical solution of μ when solving $\sum_k b_k = B$, by leveraging the monotonicity of the *Lambert W* function. ■

The complete algorithm of finding the $\mathbf{V}$ and b_k when λ_k and β_k are given is presented in Algorithm 2.

2) Update Lagrange Multipliers λ_k and β_k: Now we update the Lagrange multipliers λ_k and β_k so that (43) will be satisfied. We follow a similar step as in [20], [28], [29] with the simple gradient method. Specifically, we choose initial values of the Lagrange variables and then a standard Newton-like method is used to update the Lagrange multipliers, as

$$\lambda_k^{t+1} = \lambda_k^t + \xi^{i(n)} \nabla_1 \tag{50a}$$

$$\beta_k^{t+1} = \beta_k^t + \xi^{i(n)} \nabla_2. \tag{50b}$$

Algorithm 2 Joint Optimization of $\mathbf{V}$ and b_k for Given λ_k and β_k

1: Initialization b_k^t and set $t = 1$;
2: **repeat**
3: Obtain $\mathbf{V}^t$ by solving problem (45) with SDP;
4: Recover $\mathbf{v}^t$ from $\mathbf{V}^t$ with Gaussian randomization algorithm;
5: Update the new variable $\mathbf{V}^t$;
6: Obtain b_k^t from (46);
7: $t = t+1$;
8: **until** the objective function in (P3d') does not decrease

Algorithm 3 Joint Optimization of $\mathbf{V}$ and b_k

1: Initialize λ_k^t and β_k^t according to (43);
2: Set $t = 1$;
3: **repeat**
4: When λ_k^t and β_k^t is given, obtain $\mathbf{V}^t$ and b_k^t with Algorithm 2;
5: Update λ_k^{t+1} and β_k^{t+1} according to (50);
6: $t = t+1$;
7: **until** $\phi_k(\lambda_k^{t+1})$ and $\psi_k(\beta_k^{t+1})$ approache zero;

Here t is the iteration index, $\xi^{i(n)}$ is the step size, and ∇_1 and ∇_2 are the gradient directions for λ_k^t and β_k^t, respectively, given by

$$\nabla_1 = -\frac{\phi_k(\lambda_k)}{\phi'(\lambda_k)}, \nabla_2 = -\frac{\psi_k(\beta_k)}{\psi'(\beta_k)}.$$

We also have

$$\phi_k(\lambda_k) = \lambda_k b_k^* \log_2\left(1 + \frac{p_k f_k(\mathbf{V}^*)}{N_0 b_k^*}\right) - 1$$

$$\psi_k(\lambda_k) = \beta_k b_k^* \log_2\left(1 + \frac{p_k f_k(\mathbf{V}^*)}{N_0 b_k^*}\right) - (p_{c,k} + p_k - A_k),$$

and n is the smallest integer among {1, 2, ..., } satisfying

$$\begin{aligned}&\sum_k \left|\phi_k\left(\lambda_k^{t+1}\right)\right|^2 + \sum_k \left|\psi_k\left(\beta_k^{t+1}\right)\right|^2 \\ &\quad \leq \left(1 - \epsilon \xi^{i(n)}\right)^2 \left(\sum_k \left|\phi_k\left(\lambda_k^t\right)\right|^2 + \psi_k\left(\lambda_k^t\right)|^2\right),\end{aligned} \tag{51}$$

where $\epsilon \in (0, 1)$.

Since $(1-\epsilon\xi^{i(n)})^2$ will be a random number between [0, 1], inequality (51) will ensure that $\phi_k(\lambda_k^{t+1})$ and $\psi_k(\beta_k^{t+1})$ both go to zero, which is exactly what the optimal solution in (43) suggests. The joint optimization algorithm is summarized in Algorithm 3.

Theorem 7: Algorithm 3 will converge after a finite number of iteration steps.

Proof: Algorithm 3 is a two-layer alternating optimization algorithm. In the outer layer, the auxiliary variable λ_k and β_k are updated with a Newton-like method, the convergence of which has been proved in [30]. We only need to show that the inner layer iteration (Algorithm 2) converges, where the variables $\mathbf{V}$ and b_k are optimized.

Algorithm 4 Energy-Efficient Federated Learning

Initialize IRS phase shift matrix $\mathbf{\Theta}$ and the iteration number $t = 1$;
repeat
  Obtain f_k^t according to (13);
  Obtain p_k^t according to (17);
  Obtain b_k^t, $\mathbf{v}^t$ with Algorithm 3;
  Obtain η^t according to Theorem 3;
  Calculate the total energy consumption E^t;
  $t = t+1$;
until $\frac{|E^{t+1}-E^t|}{|E^t|} \leq \epsilon_1$ and (C1) is satisfied

Denote the objective function of (P3d') as $f(b_k, \mathbf{V})$. In the sth iteration, we have

$$f(b_k^s, \mathbf{V}^s) \overset{(a)}{\leq} f\left(b_k^s, \mathbf{V}^{s+1}\right) \overset{(b)}{\leq} f\left(b_k^{s+1}, \mathbf{V}^{s+1}\right).$$

Note that the above inequalities (a)-(b) hold true because Problems $\mathbf{V}^s$ and b_k^s are both optimally solved in each iteration s. However, we have to mention that inequality (a) will not hold strictly since we deal with the non-convex rank one constraint with the Gaussian randomization method, which may violate the monotonic improvement property of the above equation. To tackle this issue, our solution is to perform a significant number of randomization processes and select the best solution that maximizes the objective function in (P3d'). In simulations, we perform 100 Gaussian randomization and select the best $\mathbf{v}$ that achieves the maximum objective function. As a result, the inequality (a) will be guaranteed. Due to limited BS power and the finite number of IRS reflecting elements, the objective function in (P3d') is lower bounded and will converge after a finite number of steps. ∎

D. Design of the Accuracy Parameter

Finally, we optimize the accuracy parameter η. The objective function can be written as

$$f(\eta) = \frac{1}{1-\eta}\left(u + v\log\left(\frac{1}{\eta}\right)\right), \quad \eta \in (0,1), \tag{52}$$

where $u = \sum_k (p_{c,k} + p_k)\frac{S}{R}$ and $v = \sum_k \kappa D_k c_k f_k^2 \ell_1$. The optimal η can be similarly obtained as in the single user case with the bisection algorithm.

The complete algorithm for energy-efficient federated learning is presented in Algorithm 4. Note that the variables involved generally has a closed-form expression or can be obtained via simple one dimensional search with neglect-able complexity except for variable b_k and $\mathbf{v}$, which requires solving an SDP problem. Generally, solving an SDP problem with the interior method or with general CVX solvers such as MOSEK [31] incurs high complexity. According to [32, Th. 3.12], the complexity of solving an SDP problem with m constraints and an $n\times n$ variable matrix is $\mathcal{O}(\sqrt{n}\log(1/\epsilon)(mn^3 + m^2n^2 + m^3))$, where ϵ is the solution accuracy. In this problem, we have $n = N+1$ and $m = N+1$, hence the approximate complexity for solving one SDP problem would be $\mathcal{O}(\sqrt{N+1}\log(1/\epsilon)(N+1)^4)$. Suppose the iterations for Algorithm 2, Algorithm 3 and Algorithm 4 are I_2, I_3, and I_4, respectively. Then the proposed Algorithm 4 needs to solve a standard SDP problem (45) for $I_2 I_3 I_4$ times. Hence the total complexity of Algorithm 4 would be $\mathcal{O}(\sqrt{N+1}\log(1/\epsilon)I_2 I_3 I_4 (N+1)^4)$. When the number of the reflecting elements in the IRS becomes large, the total complexity would become considerably high.

VI. Low Complexity Algorithm

As analyzed before, the complexity of the proposed Algorithm 4 mainly comes from solving the SDP problem (45). To reduce the complexity, or more specifically, to reduce the complexity of getting $\mathbf{v}$ and b_k in Algorithm 2, we propose to leverage the majorization-minimization (MM) algorithm [33]. The idea is to find an easy-to-solve surrogate problem with a surrogate objective function to problem (45), and then solve this problem induced from the surrogate objective function instead of the original one. This approach can generate a sequence of sub-optimal solutions $\mathbf{v}^t$ at each iteration to approach the global optimal solution.

To proceed, we rewrite problem (45) as

$$\begin{aligned} \max_{\mathbf{v}} \quad & \sum_k \lambda_k \beta_k b_k g_k(\mathbf{v}) \\ \text{s.t.} \quad & v_{N+1} = 1; \quad |v_n| = 1, \quad \forall 1 \leq n \leq N, \end{aligned} \tag{53}$$

where $g_k(\mathbf{v}) = \log_2(1 + \frac{p_k(\mathbf{v}^H\mathbf{R}_k\mathbf{v} + \mathbf{h}_{\mathrm{d},k}^H\mathbf{h}_{\mathrm{d},k})}{N_0 b_k})$. To show the hidden convexity of $g_k(\mathbf{v})$, we have

$$g_k(\mathbf{v}) = -\log_2\left(1 - \frac{p_k\left(\mathbf{v}^H\mathbf{R}_k\mathbf{v} + \mathbf{h}_{\mathrm{d},k}^H\mathbf{h}_{\mathrm{d},k}\right)}{M_k}\right), \tag{54}$$

where $M_k = N_0 b_k + p_k(\mathbf{v}^H\mathbf{R}_k\mathbf{v} + \mathbf{h}_{\mathrm{d},k}^H\mathbf{h}_{\mathrm{d},k})$. Then $g_k(\mathbf{v}, M_k)$ is jointly convex in terms of $\{\mathbf{v}, M_k\}$ [34]. Its lower bound surrogate function is given by

$$\begin{aligned} g_k(\mathbf{v}, M_k) \geq\ & g_k\left(\mathbf{v}^t, M_k^t\right) \\ & + \frac{\partial g_k}{\partial M_k}\Big|_{M_k = M_k^t}\left(M_k - M_k^t\right) \\ & + \left(\mathbf{v} - \mathbf{v}^t\right)^H \frac{\partial g_k}{\partial \mathbf{v}}\Big|_{\mathbf{v}=\mathbf{v}^t} \\ =\ & \mathrm{const}_k^t + \tau_k^t \mathbf{v}^H\mathbf{R}_k\mathbf{v} + 2\mathbf{v}^H\mathbf{r}_k^t \triangleq \tilde{g}_k\left(\mathbf{v}|\mathbf{v}^t\right), \end{aligned} \tag{55}$$

where $\tau_k^t = -\frac{p_k^2(\mathbf{v}^{t,H}\mathbf{R}_k\mathbf{v}^t + \mathbf{h}_{\mathrm{d},k}^H\mathbf{h}_{\mathrm{d},k})}{M_k^t N_0 b_k \ln 2}$, $\mathbf{r}_k^t = \frac{p_k\mathbf{R}_k}{N_0 b_k \ln 2}\mathbf{v}^t$ and

$$\frac{\partial g_k}{\partial M_k}\Big|_{M_k = M_k^t} = -\frac{p_k\left(\mathbf{v}^{t,H}\mathbf{R}_k\mathbf{v}^t + \mathbf{h}_{\mathrm{d},k}^H\mathbf{h}_{\mathrm{d},k}\right)}{M_k^t N_0 b_k \ln 2}$$

$$\frac{\partial g_k}{\partial \mathbf{v}}\Big|_{\mathbf{v}=\mathbf{v}^t} = \frac{2p_k\mathbf{R}_k}{N_0 b_k \ln 2}\mathbf{v}^t.$$

It can be seen that $\tilde{g}_k(\mathbf{v}|\mathbf{v}^t)$ is twice differentiable and concave. Moreover, we can verify that (i) $\tilde{g}_k(\mathbf{v}^t|\mathbf{v}^t) = g_k(\mathbf{v}^t)$; (ii)$\tilde{g}_k(\mathbf{v}|\mathbf{v}^t) \leq g_k(\mathbf{v})$; and (iii) $\nabla\tilde{g}_k(\mathbf{v}|\mathbf{v}^t) = \nabla\tilde{g}_k(\mathbf{v}^t)$. Hence $\tilde{g}_k(\mathbf{v})$ is minorized at any $\mathbf{v}^n$ with a function $\tilde{g}_k(\mathbf{v}|\mathbf{v}^t)$ [33], [34]. The MM method can be used to find a

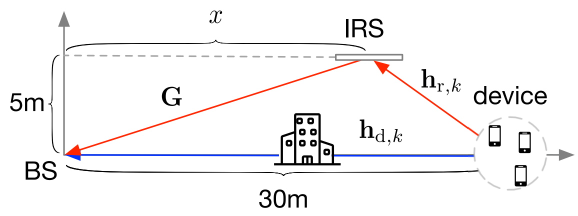

Fig. 2. Illustrate the deployment of the IRS-assisted federated learning system.

sequence of solutions to approach the global optimal solution with low complexity.

We can rewrite the objective function in (53) as $\sum_k \lambda_k \beta_k b_k \tilde{g}_k(\mathbf{v}|\mathbf{v}^t)$. Put the expression of $\tilde{g}_k(\mathbf{v}|\mathbf{v}^t)$ into the objective function and remove the constant. Accordingly, we need to solve the following problem at each iteration t.

$$\begin{aligned} &\min_{\mathbf{v}} \ \mathbf{v}^H \mathbf{R}^t \mathbf{v} - 2\mathrm{Re}\Big(\mathbf{v}^H \mathbf{r}^t\Big) \\ &\mathrm{s.t.}\ \ v_{N+1} = 1; \quad |v_n| = 1,\ \forall 1 \le n \le N, \end{aligned} \tag{56}$$

where $\mathbf{R}^t = -\sum_k \lambda_k \beta_k b_k \tau_k^t \mathbf{R}_k$ and $\mathbf{r}^t = \sum_k \lambda_k \beta_k b_k \mathbf{r}_k^t$.

Proposition 1: The objective function in (56) can be approximated by [33]:

$$\begin{aligned} &\mathbf{v}^H \mathbf{R}^t \mathbf{v} - 2\mathrm{Re}\Big(\mathbf{v}^H \mathbf{r}^t\Big) \\ &\le \mathbf{v}^H \mathbf{\Gamma}^t \mathbf{v} - 2\mathrm{Re}\Big(\mathbf{v}^H \big[\mathbf{r}^n + \big(\mathbf{\Gamma}^t - \mathbf{R}^t\big)\mathbf{v}^t\big]\Big) \\ &\quad + \mathbf{v}^{t,H}\big(\mathbf{\Gamma}^t - \mathbf{R}^t\big)\mathbf{v}^t \\ &= \rho_{\max}(\mathbf{R}^t)\mathbf{v}^H\mathbf{v} - 2\mathrm{Re}\Big(\mathbf{v}^H \tilde{\mathbf{r}}^t\Big) + \mathrm{const}', \end{aligned} \tag{57}$$

where $\lambda_{\max}\{\mathbf{R}^t\}$ is the maximum eigenvalue of matrix $\mathbf{R}^t$, $\mathbf{\Gamma}^t = \lambda_{\max}\{\mathbf{R}^t\}\mathbf{I}_{N+1}$ and $\tilde{\mathbf{r}}^t = \mathbf{r}^t + (\lambda_{\max}\{\mathbf{R}^t\}\mathbf{I}_{N+1} - \mathbf{R}^t)\mathbf{v}^t$.

To minimize the objective function in (56), we can optimize its upper bound (57). Note that $\mathbf{v}^H\mathbf{v} = N+1$ since $|v_n| = 1, \forall n$. Hence, we only need to maximize the term $2\mathrm{Re}(\mathbf{v}^H\tilde{\mathbf{r}}^t)$. This term is maximized when the phase of $\mathbf{v}$ and the phase of $\tilde{\mathbf{r}}^t$ are the same, i.e.,

$$v_i = \exp\big\{j \arg\big(\tilde{r}_i^t\big)\big\}, \quad \forall 1 \le i \le N. \tag{58}$$

It can be seen that the phase vector has a closed-form expression (58). The complexity of the proposed algorithm mainly comes from computing the eigenvalues of matrix $\mathbf{R}^t \in \mathbb{C}^{(N+1)\times(N+1)}$, which has a complexity of $\mathcal{O}((N+1)^3)$. Hence the complexity would be $\mathcal{O}(I_2 I_3 I_4 (N+1)^3)$.

VII. Simulation Study

In this section, simulation results are presented to validate the performance of the proposed IRS-assisted federated learning system. The federated learning parameters follow a similar setting as in [19], [35]. The IRS related parameters are set based on the setting in [16]. Specifically, we consider an IRS assisted communication scenario as depicted in Fig. 2. In this *x-y* plane, the IRS is located at location (20,*x*)*m*. The default value of *x* is 20m in this paper. The IoT devices are located randomly in a disk area around center (30, 0)*m* with a radius of 2m. The BS is located at the origin (0,0)m. In this section, we will change the location of the IRS and investigate the impact of such changes on the overall system performance. The channel gains are a combination of distance-dependent large-scale fading and small-scale fading. The small-scale fading is assumed to be Rayleigh fading $\mathcal{CN}(0,1)$. The large scale path loss model follows $Ad^{-\alpha}$, where $A = -30$dB is the path loss at a reference distance 1m, d is the distance between the transmitter and receiver, and α is the path loss component. The path loss components for channels $\mathbf{h}_{\mathrm{r},k}$, $\mathbf{h}_{\mathrm{d},k}$, and $\mathbf{H}$ are set to 2.2, 3.5, and 2.2, respectively. The noise power N_0 is set to $10^{-10}\,W/Hz$. The global training completion deadline is set as $T = 40$ s. For the bisection algorithm, the target accuracy is set to 10^{-5}. For Algorithm 1 and Algorithm 4, the stopping criteria is set to $\epsilon_1 = 0.01$. Each simulation result is the average of over 300 realizations.

TABLE I
Federated Learning Parameter Setting

Parameter	*Notation*	*Value*
Local sample data size	D_k	[8,12] MB
Number of CPU cycles to process one bit	c_k	30 cycles/bit
Chip energy coefficient	κ	2×10^{-28}
Training completion deadline	T_k	40s
Upload model size	S	7850 bit
IoT device static power	$p_{c,k}$	0.5 W
Maximum operating frequency	$f_{\max}$	1 GHz
Maximum transmit power	$P_{\max}$	20 W
Bandwidth	B	1 MHz
Noise power	N_0	10^{-10} W/Hz

The following two benchmark algorithms are also simulated for comparison purpose.

1) *IRS with Random Phase:* The IRS uses random phases. The detection vector $\mathbf{w}_k^t$, frequency f_k, power p_k, bandwidth b_k, and the local accuracy parameter η are optimally designed as in the proposed scheme.
2) *Without IRS:* There is only the direct channel between IoT devices and the BS. The other parameters are set as the same as in the IRS with Random Phase case.

A. Impact of the Number of Reflecting Elements N

First of all, we verify the performance of the proposed low-complexity algorithm by changing the number of reflecting elements on the IRS. In Fig. 3, we set $K = 5$, $M = 4$, and $T = 40$s and compare the energy consumption performance of different schemes. The SDP algorithm denotes Algorithm 4 where problem (45) is solved using SDP. It can be seen that with the increase of the number of reflecting elements on the IRS, the energy consumptions of the proposed low-complexity algorithm and the SDP algorithm both decrease. This is because the IRS can reconfigure the environment and help the devices to save model uploading power. A larger number of reflecting elements on the IRS generally brings a better performance. However, the processing complexity in optimizing the elements would also become quite high. Moreover, we find that the energy consumption curves for the case without IRS and IRS with random phase shift look like horizontal lines. This is straightforward as anticipated. For the case without IRS, changing the number of reflecting elements on the IRS will have no impact on the energy consumption performance. For the case IRS with random phase shift, the

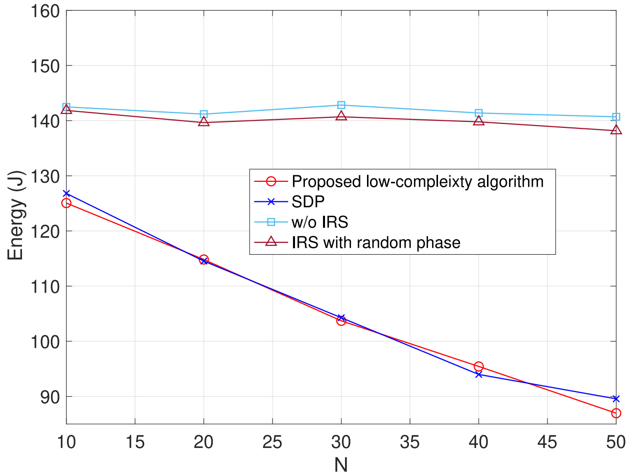

Fig. 3. Total energy consumption versus N when $K = 5$, $M = 4$, and $T = 40$.

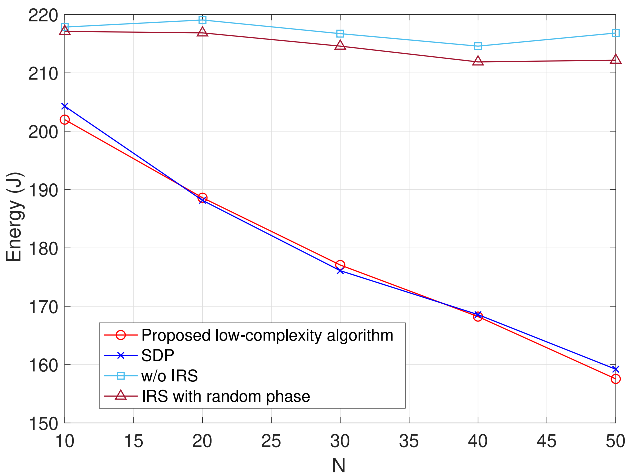

Fig. 4. Total energy consumption versus N when $K = 10$, $M = 5$ and $T = 40$.

performance does not improve significantly since the channels are not properly configured.

In Fig. 4, we perform similar experiments with $K = 10$, $M = 5$, and $T = 40$s. When the number of reflecting elements on the IRS is increased to 50, the proposed algorithm can save up to about 55 Joule compared with no IRS deployment and IRS with random phase shift. These results again demonstrate the importance of jointly optimizing resource allocation and IRS beamforming. From both Fig. 3 and Fig. 4, the performance of the proposed algorithm and the SDP algorithm achieves very similar performance, but the former has a significantly lower complexity and runs much faster. Hence, for the rest simulations, we will only consider the proposed low-complexity algorithm.

B. Convergence Behavior

In this section, we investigate the convergence of the proposed low-complexity algorithm. The convergence behavior of the general federated learning scheme is plotted in Fig. 5. It can be seen that after 2-3 iterations, the energy consumption decreases to a low value. However, the completion time might violate the training deadline constraint. After

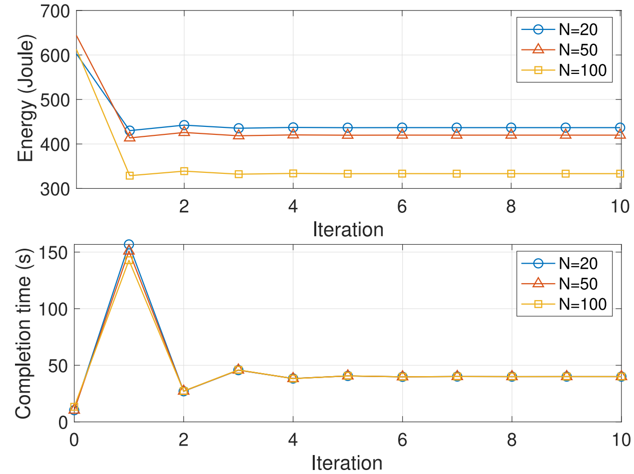

Fig. 5. Convergence of the first device in a multiuser federated learning system with $K = 20$, $M = 4$, and $T = 40$.

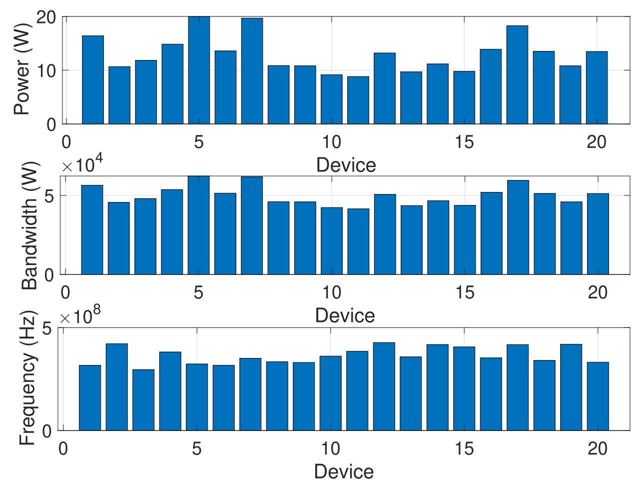

Fig. 6. Power, bandwidth, and frequency allocation for different devices with $K = 20$, $M = 4$, $N = 20$, and $T = 40$.

several fine-tuning iterations, we can obtain a feasible solution that minimizes the energy consumption while also satisfying the completion time constraint.

We also change the number of the reflecting elements on the IRS. We find that when $N = 50$, the device saves more energy than the case when $N = 20$. Despite that, convergence of the training process does not change much when N is varied.

C. Energy and Time Consumption of Each Device

The power, bandwidth, and frequency allocation parameters for different devices are presented in Fig. 6. In this simulation, the devices are located very close to each other. Their operating frequency, power, and bandwidth seem not differ too much. The total energy consumption and time consumption over the entire training process is shown in Fig. 7. It can be seen that all the devices share the same latency, which is exactly $T = 40$s, while the consumed energy differs. With our parameter setting, we also find that local training nearly does not consume much energy but it accounts for almost 99% of latency. On the contrary, model uploading takes a lot of energy while it nearly takes no time. This setting is reasonable since in practice the

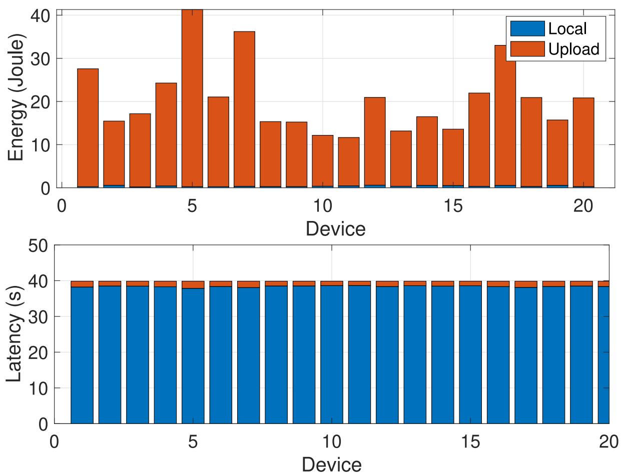

Fig. 7. Total energy consumption and latency of local model training and model uploading for different devices with $K = 20$, $M = 4$, and $T = 40$.

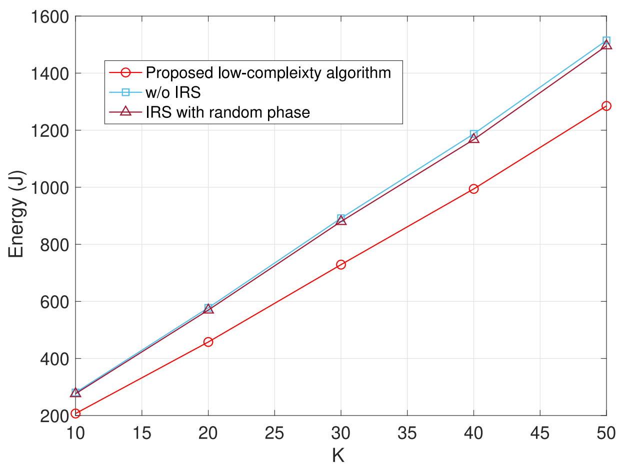

Fig. 8. Total energy consumption versus K with $M = 4$, $N = 40$, and $T = 40$.

model is trained locally by each device. The training usually takes several rounds which take time. Moreover, the device works on the lowest possible frequency, which further slows down the completion time. On the other hand, the devices are battery powered, the model update process consumes most of the energy. In this case, the deployed IRS can work as a passive, enhanced channel, which helps the devices to save their battery power.

D. Impact of the Number of Devices K

We investigate the impact of the number of devices in Fig. 8. We find that the energy consumption generally increases linearly with the number of devices involved. This is because in the multiuser system, only the bandwidth and the IRS reflecting elements are optimized jointly. Each device selects its own operating frequency and power. With increased number of devices, the proposed algorithm saves more energy than the two baseline algorithms. Moreover, the performance gap becomes larger as K is increased. This result demonstrates the advantages of the proposed algorithm in a communication system where the number of IoT devices is large.

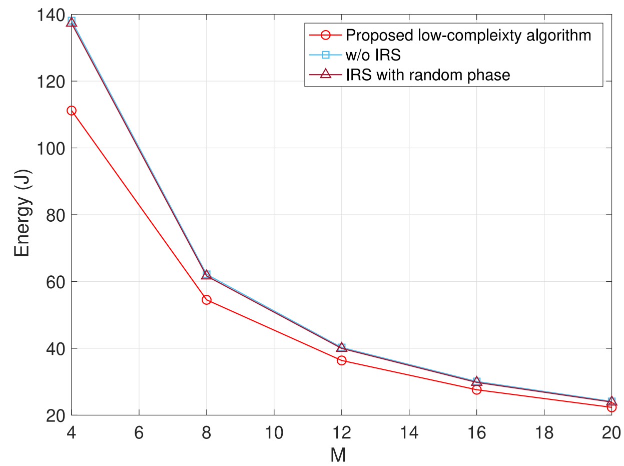

Fig. 9. Total energy consumption versus M with $K = 5$, $N = 20$, and $T = 40$.

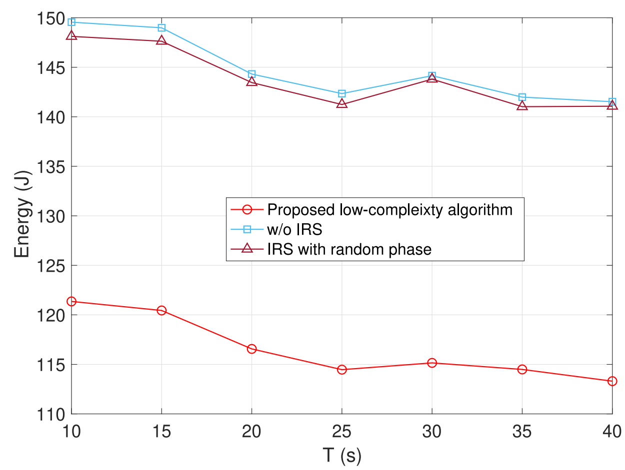

Fig. 10. Total energy consumption versus T with $K = 5$, $N = 20$, and $M = 4$.

E. Impact of the Number of Antennas M on the BS

Fig. 9 shows the impact of the number of BS antennas on energy saving of the federated learning system. As can be seen, with more receiving antennas on the BS, the system energy consumption can be greatly reduced. This is because the antennas on the BS provide additional multiplexing gain at the receiver so that each IoT device can reduce their transmit power for model uploading. Moreover, the performance gap between the proposed algorithm and the two benchmark algorithms will gradually vanish with increased M. This motivates us to deploy an IRS with a larger number of reflecting elements, i.e., $N > M$, to harvest the reconfigured channel gain provided by the IRS.

F. Impact of the Task Completion Time T

Fig. 10 shows the impact of task completion time T on energy saving of the federated learning system. It can be seen that the energy consumption slightly decreases with the increase of the completion time. This is because the devices always work on the lowest frequency to save energy and satisfy the task completion time. Moreover, in our setting the local computing takes a lot of time but only accounts for a small portion of energy consumption, while model uploading takes

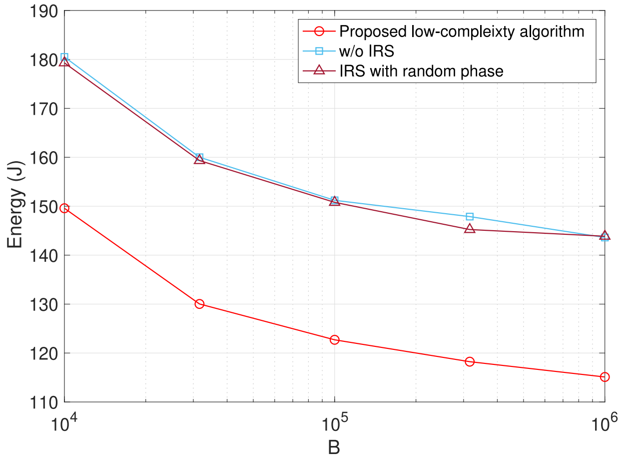

Fig. 11. Total energy consumption versus bandwidth B with $K = 5$, $N = 20$, and $M = 4$.

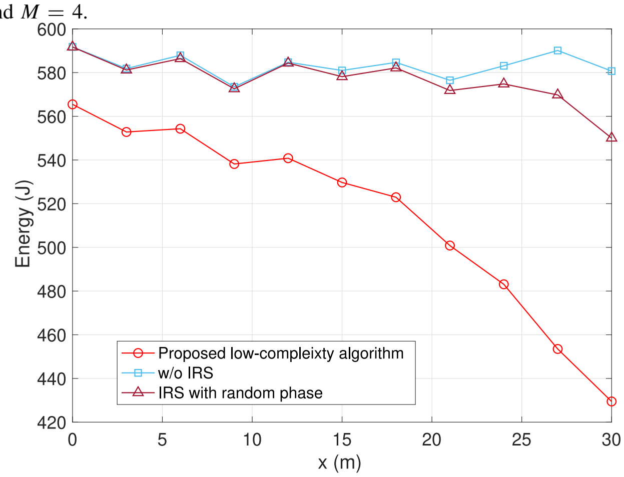

Fig. 12. Energy consumption versus the location of the IRS when $N = 20$, $K = 20$, $M = 4$, and $T = 40$.

little time but consumes a lot of energy. In other words, the total energy consumption of the proposed federated learning system is insensitive to the task completion time.

G. Impact of the Bandwidth Constraint B

Fig. 11 shows the impact of the communication bandwidth on the system energy consumption. With the increase of the available bandwidth, each IoT device can reduce their transmit power or their uploading time to upload the same model. Hence the total energy consumption can be saved. As can be seen, the absolute value of the slope of these curves gradually goes to zero, which suggests that the impact of bandwidth is diminishing in the high bandwidth region. In other words, when the available bandwidth is large enough, the other communication/computing factors will become the major factor(s) that prevent the reduction of energy consumption.

H. Impact of the IRS Location and Path Loss on the Reflecting Channel

The impact of the IRS location on the system energy consumption is presented in Fig. 12, where x measures the distance between the IRS and BS. When the IRS is close to the IoT devices, the energy saving will be significant. The impact

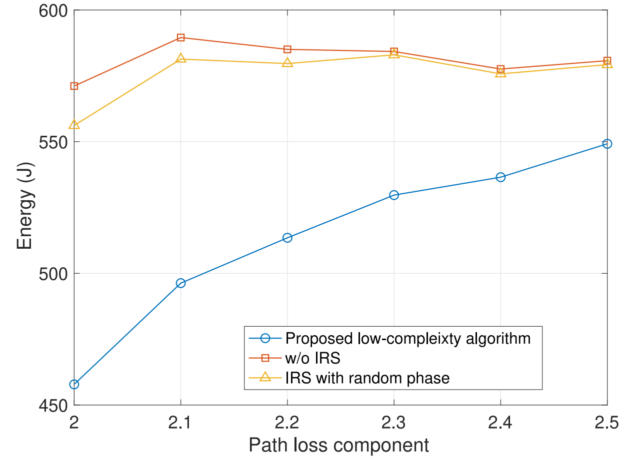

Fig. 13. Energy consumption versus the path loss of the reflecting channel.

of the location of the IRS depends on the IRS reflected channel fading. In practice, the location of the IRS should be properly selected to reap the maximum benefit of the IRS technology. Similarly, the energy consumption versus the path loss of the reflecting channel is shown in Fig. 13. The default setting on the reflected channel is $\alpha = 2.2$ for $\mathbf{h}_{\mathrm{r},k}$ and $\mathbf{H}$. Now we change the value of the path loss from 2 to 2.5. We find that when the path loss on the reflected channel becomes larger, the energy saving becomes less. This is easy to explain. When the path loss on the reflected channel becomes larger, the channel enhancement effect of the IRS will become weaker. In the extreme case when the path loss on the reflected channel is infinitely large, i.e., the reflected channel is blocked, the deployment of IRS will make on difference.

Our simulation is based on the assumption that the repeated model uploading accounts for the major energy consumption of the federated learning system. In some systems, the local computing may take up the major energy consumption compared with the communication process. Different system factors such as local data size, model accuracy level, environment noise power level, and the CPU processing capability may have various effects on the system trade-offs: 1) between task completion time and the energy consumption and 2) energy consumption caused by communication and computation. The proposed algorithm provides a low complexity solution to explore these trade-offs.

VIII. Conclusion

In this paper, we considered an energy-efficient federated learning framework where devices uploads their locally trained models when assisted by an IRS. In this framework, an energy minimization problem was considered. We proposed an efficient parameter optimization algorithm to jointly optimize system parameters, such as the operating frequency of each device, transmit power, bandwidth, the IRS phases, and the local accuracy parameter. The proposed low-complexity algorithm can reasonably manage the energy resources by balancing the communication and local training costs. We have conducted extensive experiments to shed insight on the benefits on the use of IRS in federated learning systems.

References

[1] K. B. Letaief, W. Chen, Y. Shi, J. Zhang, and Y.-J. A. Zhang, "The roadmap to 6G: AI empowered wireless networks," *IEEE Commun.*, vol. 57, no. 8, pp. 84–90, Aug. 2019.

[2] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. y Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. 20th Int. Conf. Artif. Intell. Stat.*, Apr. 2017, pp. 1273–1282.

[3] J. Konečnỳ, H. B. McMahan, F. X. Yu, P. Richtárik, A. T. Suresh, and D. Bacon, "Federated learning: Strategies for improving communication efficiency," Oct. 2016, *arXiv:1610.05492*.

[4] T. Li, A. K. Sahu, A. Talwalkar, and V. Smith, "Federated learning: Challenges, methods, and future directions," *IEEE Signal Process.*, vol. 37, no. 3, pp. 50–60, May 2020.

[5] S. Zargari, A. Khalili, and R. Zhang, "Energy efficiency maximization via joint active and passive beamforming design for multiuser MISO IRS-aided SWIPT," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 557–561, Mar. 2021.

[6] Q. Wu and 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.

[7] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Trans. Commun.*, vol. 69, no. 5, pp. 3313–3351, May 2021.

[8] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun.*, vol. 58, no. 1, pp. 106–112, Jan. 2020.

[9] M. D. Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[10] E. F. Kuester, M. A. Mohamed, M. Piket-May, and C. L. Holloway, "Averaged transition conditions for electromagnetic fields at a metafilm," *IEEE Trans. Antennas Propag.*, vol. 51, no. 10, pp. 2641–2651, Oct. 2003.

[11] V. Arun and H. Balakrishnan, "RFocus: Beamforming using thousands of passive antennas," in *Proc. USENIX NSDI*, Santa Clara, CA, USA, Feb. 2020, pp. 1047–1061.

[12] Y. Yang, B. Zheng, S. Zhang, and R. Zhang, "Intelligent reflecting surface meets OFDM: Protocol design and rate maximization," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4522–4535, Jul. 2020.

[13] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019.

[14] X. Yu, D. Xu, Y. Sun, D. W. K. Ng, and R. Schober, "Robust and secure wireless communications via intelligent reflecting surfaces," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2637–2652, Nov. 2020.

[15] C. Pan *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020.

[16] Z. Chu, P. Xiao, M. Shojafar, D. Mi, J. Mao, and W. Hao, "Intelligent reflecting surface assisted mobile edge computing for Internet of Things," *IEEE Wireless Commun. Lett.*, vol. 10, no. 3, pp. 619–623, Mar. 2021.

[17] K. Yang, Y. Shi, Y. Zhou, Z. Yang, L. Fu, and W. Chen, "Federated machine learning for intelligent IoT via reconfigurable intelligent surface," *IEEE Netw.*, vol. 34, no. 5, pp. 16–22, Sep./Oct. 2020.

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

[19] N. H. Tran, W. Bao, A. Zomaya, M. N. Nguyen, and C. S. Hong, "Federated learning over wireless networks: Optimization model design and analysis," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2019, pp. 1387–1395.

[20] T. Bai, C. Pan, Y. Deng, M. Elkashlan, A. Nallanathan, and L. Hanzo, "Latency minimization for intelligent reflecting surface aided mobile edge computing," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2666–2682, Nov. 2020.

[21] Z. Wang *et al.*, "Federated learning via intelligent reflecting surface," 2020, *arXiv:2011.05051*.

[22] T. T. Vu, D. T. Ngo, N. H. Tran, H. Q. Ngo, M. N. Dao, and R. H. Middleton, "Cell-free massive mimo for wireless federated learning," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6377–6392, Oct. 2020.

[23] J. Konečnỳ, Z. Qu, and P. Richtárik, "Semi-stochastic coordinate descent," *Optim. Methods Softw.*, vol. 32, no. 5, pp. 993–1005, Mar. 2017.

[24] C. Ma *et al.*, "Distributed optimization with arbitrary local solvers," *Optim. Methods Softw.*, vol. 32, no. 4, pp. 813–848, 2017.

[25] Y. Wang, M. Sheng, X. Wang, L. Wang, and J. Li, "Mobile-edge computing: Partial computation offloading using dynamic voltage scaling," *IEEE Trans. Commun.*, vol. 64, no. 10, pp. 4268–4282, Oct. 2016.

[26] J. F. Epperson, *An Introduction to Numerical Methods and Analysis.* Hoboken, NJ, USA: Wiley, 2021.

[27] M. Grant, S. Boyd, and Y. Ye, *CVX Users' Guide*, CVX Res., Austin, TX, USA, 2009. [Online]. Available: http://www.stanford.edu/~boyd/software.html

[28] Y. Jong, *An Efficient Global Optimization Algorithm for Nonlinear Sum-of-Ratios Problem*. Pyongyang, North Korea: Univ. Sci., May 2012.

[29] S. He, Y. Huang, L. Yang, and B. Ottersten, "Coordinated multicell multiuser precoding for maximizing weighted sum energy efficiency," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 741–751, Feb. 2014.

[30] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[31] "Using Mosek with CVX." CVX Research. 2012. [Online]. Available: http://cvxr.com/cvx/doc/mosek.html

[32] I. Pólik and T. Terlaky, "Interior point methods for nonlinear optimization," in *Nonlinear Optimization*. Heidelberg, Germany: Springer, 2010, pp. 215–276.

[33] Y. Sun, P. Babu, and D. P. Palomar, "Majorization-minimization algorithms in signal processing, communications, and machine learning," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 794–816, Feb. 2017.

[34] G. Zhou, C. Pan, H. Ren, K. Wang, and A. Nallanathan, "Intelligent reflecting surface aided multigroup multicast MISO communication systems," *IEEE Trans. Signal Process.*, vol. 68, pp. 3236–3251, Apr. 2020.

[35] R. Jin, X. He, and H. Dai, "On the design of communication efficient federated learning over wireless networks," Jun. 2020, *arXiv:2004.07351*.

Ticao Zhang received the B.E. and M.S. degrees from the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, in 2014 and 2017, respectively. He is currently pursuing the Ph.D. degree in electrical and computer engineering with Auburn University. His research interests include video coding and communications, machine learning, and optimization and design of wireless multimedia networks. His paper was featured as IEEE ACCESS Journal's "Article of the Week" in April 2020.

Shiwen Mao (Fellow, IEEE) received the Ph.D. degree in electrical engineering from Polytechnic University, Brooklyn, NY, USA, in 2004. He is currently a Professor and the Earle C. Williams Eminent Scholar of Electrical and Computer Engineering with Auburn University, Auburn, AL, USA. His research interests include wireless networks, multimedia communications, and smart grid. He is a co-recipient of the 2021 IEEE INTERNET OF THINGS JOURNAL Best Paper Award, the 2021 IEEE Communications Society Outstanding Paper Award, the IEEE Vehicular Technology Society 2020 Jack Neubauer Memorial Award, the IEEE ComSoc MMTC 2018 Best Journal Paper Award and the 2017 Best Conference Paper Award, the Best Demo Award of IEEE SECON 2017, the Best Paper Awards of IEEE GLOBECOM 2019, 2016, and 2015, IEEE WCNC 2015, and IEEE ICC 2013, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is a Distinguished Lecturer of IEEE Communications Society from 2021 to 2022 and IEEE Council of RFID from 2021 to 2022, and a Distinguished Lecturer from 2014 to 2018 and a Distinguished Speaker of IEEE VEHICULAR TECHNOLOGY SOCIETY from 2018 to 2021. He is on the Editorial Board of IEEE/CIC *China Communications*, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE INTERNET OF THINGS JOURNAL, IEEE OPEN JOURNAL OF THE COMMUNICATIONS SOCIETY, *ACM GetMobile*, IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, IEEE TRANSACTIONS ON NETWORK SCIENCE AND ENGINEERING, IEEE TRANSACTIONS ON MOBILE COMPUTING, IEEE MULTIMEDIA, IEEE NETWORK, and IEEE NETWORKING LETTERS. He is a member of ACM.