Reconfigurable Intelligent Surface-Enabled Federated Learning for Power-Constrained Devices

Quang Nhat Le®, Graduate Student Member, IEEE, Lina Bariah®, Senior Member, IEEE, Octavia A. Dobre®, Fellow, IEEE, and Sami Muhaidat®, Senior Member, IEEE

Abstract—Federated learning (FL) has recently emerged as a novel technique for training shared machine learning models in a distributed fashion while preserving data privacy. However, the application of FL in wireless networks poses a unique challenge on the mobile users (MUs)’ battery lifetime. In this letter, we aim to apply reconfigurable intelligent surface (RIS)-aided wireless power transfer to facilitate sustainable FL-based wireless networks. Our objective is to minimize the total transmit power of participating MUs by jointly optimizing the transmission time, power control, and the RIS’s phase shifts. Numerical results demonstrate that the total transmit power is minimized while satisfying the requirements of both minimum harvested energy and transmission data rate.

Index Terms—Energy harvesting, federated learning, optimization, reconfigurable intelligent surface.

I. INTRODUCTION

W ITH the recent industrial and research activities devoted to the sixth generation (6G) wireless networks, the uniqueness of the future 6G networks—compared to previous wireless network generations—lies in the realization of ubiquitous intelligence. In this sense, the artificial intelligence (AI) will be exploited to orchestrate wireless networks from the core to the edge [1]. To this end, machine learning (ML), which is a subfield of AI, is anticipated to be an indispensable tool in future 6G networks. ML will operate on the data collected from all network segments in order to enable smart resource management, access control, multi-layer communication, etc. The evolution of the ML paradigm was initially fueled by the huge amount of data generated from a wide range of emerging applications, including smart cities, autonomous vehicles, haptics, and augmented reality services, to name a few. Accordingly, the ML algorithms are key to leverage such data with the aim of training wireless networks, and therefore, enabling truly self-optimized and self-organized networks. Since their initial development, classical ML algorithms demonstrated a superior performance in handling complicated tasks pertaining to channel estimation, spectrum sensing, resource allocation, etc. Nevertheless, given that the underlying principle of ML is the collection of raw data generated at end-devices and stored at a centralized server for model training purposes, a number of concerning issues have been flagged. First, in cloud-based ML algorithms, users privacy is compromised due to the exchange of their local datasets, exposing participating mobile users (MUs) to potential security attacks. Second, the long propagation delay in centralized ML algorithms limits their applications in real-time scenarios. Finally, the centralized ML paradigm suffers from the increased network overhead, rendering it unsuitable for power-constrained MUs [2].

Recently, federated learning (FL) has been identified as an efficient decentralized learning mechanism, in which improved data privacy, reduced latency, and network overhead can be achieved [3], [4]. FL is a collaborative learning mechanism, in which the on-board computing and storage capabilities, in addition to the local datasets of participating MUs (clients), are leveraged to perform local model training. The locally trained model updates are then shared with a cloud-based server for aggregation and global model evaluation. These steps are repeated until a desired level of accuracy is attained. Given that less information is required to be shared with the central server, compared to classical cloud-centric ML algorithms, FL features enhanced user privacy and offers a better utilization for the network resources. Inspired by its promising potentials, extensive research attempts have been initiated to investigate the advantages of the interplay of FL and other enabling technologies, including reconfigurable intelligent surface (RIS), blockchain, aerial/satellite communications, and mobile edge computing [3], [4], [5], [6], [7], [8].

Despite the several advantages of FL, only few attempts have explored the problem of executing energy-consuming computing and communication tasks at MUs with limited energy budget. In [9], the problem of joint delay and energy minimization in an IoT network was investigated by using a three-tier offloading scheme. The authors in [10] minimized the overall energy consumption at participating MUs using the non-orthogonal multiple access (NOMA) scheme at the uplink transmission. This was accomplished by optimizing the number of iterations a client is requested to update its local model, under a particular global model accuracy threshold. On the other hand, the work in [11], [12], [13], and [14] considered the adoption of the wireless power transfer (WPT) concept as a mean to supply power-limited MUs and enable them to participate in the training process. In particular, in [11], a hybrid radio frequency (RF)/visible light communication (VLC) scenario was considered, in which the VLC link...
over the downlink is leveraged for energy harvesting (EH) purposes, while the RF is used for the local model updates transmission. From a similar perspective, the authors in [12] applied the time-switching paradigm of WPT to an FL system, in which they studied the trade-off between learning and WPT, and further optimized the MU clock frequency for improved utilization of harvested energy for model evaluation. From a similar point of view, the authors in [14] minimized the mean squared error (MSE) by optimizing the beamforming and consumed energy in a vehicular energy-limited network. The authors in [13] leveraged RIS for WPT and developed an improved local model updates transmission. In particular, they formulated an optimization problem to obtain the RIS phase shift (PS) vector that minimizes the MSE.

Unlike the works in [10], [11], [12], and [14], which considered the application of WPT in FL, motivated by the intertwined benefits of RIS and FL paradigms, our work considers leveraging RIS for WPT purposes, to enable power-constrained MUs meet the computing and communication requirements imposed by the FL process. Instead of focusing on minimizing the MSE, as in [13], our proposed framework aims to minimize the total transmit power of the participating MUs, while satisfying particular computing and communication requirements. This is achieved by jointly optimizing the RIS and MUs operational parameters.

**Notation:** $X^T$ denotes the transpose of a matrix $X$, $\mathbb{E}[*]$ represents the expectation operation, $|\cdot|$ represents the absolute value of a complex scalar, and $\Re{\cdot}$ returns the real part of an argument.

**II. SYSTEM MODEL**

We consider a FL model, in which the downlink transmission is utilized to recharge $K$ EH-enabled MUs, while the uplink is dedicated for communicating the FL local model parameters to a base station (BS), as depicted in Fig. 1. During the downlink transmission, a BS$_1$ communicates with the $K$ MUs through the assistance of an RIS, comprising $N$ reflecting elements (REs). This is motivated by the assumption that the direct links between BS$_1$ and the $K$ devices are unavailable. Therefore, the RIS is exploited to extend the signal coverage and to enhance the received signal strength, and thus, to enhance the harvested energy [15]. Without loss of generality, we assume that each node in the network is equipped with a single antenna.

**A. Wireless Power Transfer (WPT) Model**

In the underlying system model, BS$_1$ sends the RF signal, $x_E$ with $\mathbb{E}\{|x_E|^2\} = 1$, to the $K$ MUs through the RIS. The average transmit power at BS$_1$ is represented by $P_{BS_1}$. The baseband signal received at the $k$-th MU can be written as

$$y_{BS_1,k} = \sqrt{P_{BS_1}} h_k^T \Phi g x_E + n_k = \sqrt{P_{BS_1}} \psi^T \text{diag}(g^H) h_k x_E + n_k = \sqrt{P_{BS_1}} q_k(\psi) x_E + n_k,$$

where $g \in \mathbb{C}^{N \times 1}$ and $h_k \in \mathbb{C}^{1 \times N}$ denote the channels from BS$_1$ to RIS, and from RIS to the $k$-th MU, respectively. $\Phi \in \mathbb{C}^{N \times N}$ represents the PS matrix of RIS, which can be written as: $\Phi \triangleq \text{diag}(e^{j\theta_1}, e^{j\theta_2}, \ldots, e^{j\theta_N})$, where $\theta_n \in [0, 2\pi)$ denotes the PS of the $n$-th RE on the RIS. Further, $\Phi$ can be rewritten as $\Phi = \text{diag}(\psi_1, \psi_2, \ldots, \psi_N)$, with $|\psi_n| = 1$, $\forall n \in N$. $\psi \triangleq \psi_1, \psi_2, \ldots, \psi_N)^T$ and $q_k(\psi) \triangleq \psi^T \text{diag}(g^H) h_k$. $n_k \sim \mathcal{CN}(0, \sigma^2)$ represents the additive white gaussian noise (AWGN) over the downlink transmission.

Assuming all $K$ MUs are equipped with EH devices, the harvested power at the $k$-th MU can be written as

$$P_{EH} = \eta P_{BS_1} |q_k(\psi)|^2,$$

where $\eta$ accounts for the energy conversion efficiency. In this work, we consider that the harvested power is the only source of power. In particular, the harvested power is divided into two parts, namely $\mu P_{EH}$ and $(1-\mu) P_{EH}$, which are dedicated for local model transmission and computation, respectively, with $\mu$ denoting the power splitting factor.

**B. Distributed FL Model**

Consider that all $K$ MUs are selected by BS$_2$ to perform a particular on-device distributed FL task, which aims at optimizing the model parameter $z$ that minimizes the local loss function $f_k(z)$:

$$\text{minimize}_{z \in \mathbb{R}^d} \quad f_k(z) = \frac{1}{S_k} \sum_{i \in D_k} f_i(z),$$

where $S_k$ denotes the size of the local data set $D_k$ of device $k$ and $f_i(z)$ represents the loss function associated with the data pair $i$. Without loss of generality, we assume that all local data sets have uniform size, i.e., $S_k = S$. To maintain higher spectral efficiency and reduce the total number of communication rounds between BS$_2$ and devices, a model averaging scheme [16] is utilized. In particular, BS$_2$ broadcasts the global model update $z$ to the selected $K$ MUs. At each MU, a local update algorithm is executed to generate the local updated model $\hat{z}_k$, relying on the local data set and the received global model. Then, the selected MUs send $K$ weighted local models, which are aggregated at BS$_2$ to compute the updated global model, as $z = \xi \left( \sum_{k \in K} \phi_k(\hat{z}_k) \right)$, where $\xi$ denotes the post-processing function at BS$_2$, while $\phi_k$ represents the pre-processing function at the $k$-th MU. Since the harvested energy from the RF signal constitutes the only source of energy at the K MUs, we assume that all local model parameters at each MU are transmitted over a single transmission period. Therefore, by setting $x_k = \phi_k(z_k)$ to represent the transmitted signal at the $k$-th MU, the target function to be estimated at BS$_2$ can be written as $\chi = \sum_{k \in K} x_k$. Note that the central processing unit (CPU) energy
consumed to process all data at the k-th MU over a single local iteration can be expressed as [17]

$$E_{cp}^{(k)} = \frac{\nu_k c_k S}{2} \omega_k^2,$$

where $\nu_k$ denotes the effective capacitance coefficient of the computing chipset at the k-th MU, and $c_k$ represents the number of CPU cycles required to process one sample data at the k-th MU. Also, $\omega_k$ accounts for the k-th MU’s CPU cycle frequency. The computation time for one local iteration is $T_{cp,k} = c_k S/\omega_k$. It is worth highlighting that the energy required for model transmission at the k-th MU can be evaluated as [18]

$$E_{cm}^{(k)} = P_{T,k} T_{cm,k},$$

where $P_{T,k}$ and $T_{cm,k}$ denote the transmission power and the transmission time at the k-th MU, respectively. Considering that the uplink channel between the k-th MU and the second BS is $T_k$, which follows the Rayleigh distribution, and based on the time division multiple access (TDMA), the received signal at BS2 for the k-th MU is given by

$$y_{BS2,k} = \sqrt{P_{T,k}} T_k x_k + n_{BS2},$$

where $n_{BS2} \sim CN(0, \sigma^2)$ represents the AWGN over the uplink transmission. At BS2, the target function is estimated by utilizing aggregation beamforming, $\nu$, and thus, the achievable rate of the k-th MU at BS2 can be evaluated as

$$R_k = \log_2 \left(1 + \frac{P_{T,k} |\nu K_k|^2}{\sigma^2} \right).$$

III. PROBLEM FORMULATION

Let us define $P_T \triangleq [P_{T,1}, P_{T,2}, \ldots, P_{T,K}]^T$, $T_{cm} \triangleq [T_{cm,1}, T_{cm,2}, \ldots, T_{cm,K}]^T$, and $\omega \triangleq [\omega_1, \omega_2, \ldots, \omega_K]^T$. With $\delta = 1 - \mu$, the optimization problem is formulated as

$$\min_{P_{T}, \omega, \psi, k \in \mathcal{K}} \sum_{k \in \mathcal{K}} P_{T,k}$$

s.t. $E_{cp}^{(k)}(\omega_k) \leq \delta P_{EH}(\psi) \tau, \forall k \in \mathcal{K},$ (8b)

$E_{cm}^{(k)}(P_{T,k}, T_{cm,k}) \leq P_{EH}(\psi) \tau, \forall k \in \mathcal{K},$ (8c)

$T_{cm,k} B R_k(P_{T,k}, v) \geq \Theta_k, \forall k \in \mathcal{K},$ (8d)

$M_k T_{cp,k} (\omega_k) + T_{cm,k} = \tau, \forall k \in \mathcal{K},$ (8e)

$\omega_{\min} \leq \omega_k \leq \omega_{\max}, \forall k \in \mathcal{K},$ (8f)

$|\nu| = 1, \forall n \in \mathcal{N},$ (8h)

$|\psi_n| = 1, \forall n \in \mathcal{N}.$ (8i)

In the above problem, (8b) and (8c) imply that the energy consumed by local model computation and communication at the k-th MU should be less than its corresponding harvested energy. Eq. (8d) ensures that the transmission rate of the k-th MU is greater than or equal to the required data size $\Theta_k$. Constraint (8e) ensures that each global round should be finished within the time frame $\tau$, where $M_k$ denotes the required number of local iterations at the k-th MU. The CPU frequency of the k-th MU is specified in (8f), where $\omega_{\min}$ and $\omega_{\max}$ denote the minimum and maximum CPU frequency of the k-th MU, respectively. Finally, (8h) and (8i) imply that the aggregation beamforming has unit power and unit-modulus constraint at the RIS, respectively. It is worth noting that problem (8) is non-convex and challenging to solve, due to the non-convexity nature of constraints (8c)-(8e), (8h), (8i) and the coupling of multiple optimization variables.

IV. PROPOSED SOLUTION

Assuming an iterative algorithm based on the inner approximation (IA) framework, let $(P_{T}^{(\kappa)}, \psi^{(\kappa)}, T_{cm}^{(\kappa)}, \nu^{(\kappa)})$ be the feasible point for (8) that is obtained from the $(\kappa - 1)$-th round. In this section, an alternating algorithm is proposed to solve (8), i.e., at the iteration $(\kappa + 1)$ solving (8) to find the optimal solutions $P_{T}^{*} := P_{T}^{(\kappa + 1)}, \psi^{*} := \psi^{(\kappa + 1)}, T_{cm}^{*} := T_{cm}^{(\kappa + 1)}, \mu^{*} := \mu^{(\kappa + 1)}$ for given $(P_{T}^{(\kappa)}, \psi^{(\kappa)}, T_{cm}^{(\kappa)}, \nu^{(\kappa)})$, then finding the optimal solution $\nu^{*} := \nu^{(\kappa + 1)}$ for given $(P_{T}^{(\kappa + 1)}, \psi^{(\kappa)}, T_{cm}^{(\kappa + 1)}, \nu^{(\kappa)})$, followed by finding the optimal solution $\psi^{*} := \psi^{(\kappa + 1)}$ for given $(P_{T}^{(\kappa + 1)}, \nu^{(\kappa)}, T_{cm}^{(\kappa + 1)}, \omega^{(\kappa + 1)}), \mu^{(\kappa + 1)})$.

A. Transmit Power Optimization

At iteration $(\kappa + 1)$, problem (8) for given $(P_{T}^{(\kappa)}, \psi^{(\kappa)}, T_{cm}^{(\kappa)}, \nu^{(\kappa)})$ can be expressed as

$$\min_{P_{T}, \omega, \psi, k \in \mathcal{K}} \sum_{k \in \mathcal{K}} P_{T,k}$$

s.t. $0.5 \left( T_{cm,k}^{(\kappa)} P_{T,k}^{2} + P_{T,k}^{(\kappa)} T_{cm,k}^{2} \right) \leq \mu \eta P_{bs1} \left| q_k \left( \nu^{(\kappa)} \right) \right|^2 \tau,$

where $P_{T,k} T_{cm,k} \leq 0.5 \left( T_{cm,k}^{(\kappa)} P_{T,k}^{2} + P_{T,k}^{(\kappa)} T_{cm,k}^{2} \right), \forall k \in \mathcal{K}$.

Next, (8d) is rewritten as follows:

$$T_{cm,k} \ln \left(1 + \frac{|\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}}{\sigma^2} \right) \left(2 - \frac{T_{cm,k}^{(\kappa)}}{T_{cm,k}} \right)$$

$$+ \frac{T_{cm,k}^{(\kappa)} |\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}}{\sigma^2 + |\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}} \ln \left(1 - \frac{T_{cm,k}^{(\kappa)} P_{T,k}^{(\kappa)}}{\Theta_k} \right)$$

$$\geq \Theta_k \ln 2 B, \forall k \in \mathcal{K},$$

which is equivalent to

$$T_{cm,k} \ln \left(1 + \frac{|\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}}{\sigma^2} \right)$$

$$\times \left(2 - T_{cm,k}^{(\kappa)} P_{T,k}^{(\kappa)} \right) + \frac{T_{cm,k}^{(\kappa)} |\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}}{\sigma^2 + |\nu^{(\kappa)} Y_k|^2 P_{T,k}^{(\kappa)}}$$

$$\times \left(1 - P_{T,k}^{(\kappa)} P_{T,k}^{(\kappa)} \right) \geq \Theta_k \ln 2 \frac{B}{P_{T,k}^{(\kappa)}}, \forall k \in \mathcal{K},$$

$$P_{T,k}^{(\kappa)} P_{T,k}^{(\kappa)} \geq 1, \forall k \in \mathcal{K}.$$
where $\tilde{T}_{cm} \triangleq \begin{bmatrix} \tilde{T}_{cm,1}, \tilde{T}_{cm,2}, \ldots, \tilde{T}_{cm,K} \end{bmatrix}^T$ and $\tilde{P}_T \triangleq \begin{bmatrix} \tilde{P}_{T,1}, \tilde{P}_{T,2}, \ldots, \tilde{P}_{T,K} \end{bmatrix}^T$ are new introduced variables, with $\tilde{T}_{cm,k} \geq \frac{1}{\eta P_{BS_i}}$ and $\tilde{P}_{T,k} \geq \frac{1}{\eta P_{BS_i}}$, $\forall k \in K$. Constraints (12b) and (12c) can be further expressed as the following second-order cone (SOC) constraints, for $\forall k \in K$,
\[
0.25 \left(T_{cm,k} + \tilde{T}_{cm,k}\right)^2 \geq 0.25 \left(T_{cm,k} - \tilde{T}_{cm,k}\right)^2 + 1, \quad (13)
\]
\[
0.25 \left(P_{T,k} + \tilde{P}_{T,k}\right)^2 \geq 0.25 \left(P_{T,k} - \tilde{P}_{T,k}\right)^2 + 1. \quad (14)
\]

Similarly, constraint (8e) can be equivalently rewritten as
\[
\text{(8e)} \iff \begin{cases} M_{ck}\tilde{\omega}_k + T_{cm,k} = \tau, \\ 0.25 \left(\omega_k + \tilde{\omega}_k\right)^2 \geq 0.25 \left(\omega_k - \tilde{\omega}_k\right)^2 + 1, \end{cases} \quad (15a)
\]
where $\tilde{\omega} \triangleq \begin{bmatrix} \tilde{\omega}_1, \tilde{\omega}_2, \ldots, \tilde{\omega}_K \end{bmatrix}^T$ is a new introduced variable, with $\tilde{\omega}_k \geq \frac{1}{\eta P_{BS_i}}$, $\forall k \in K$.

The equivalent convex problem of (9) for iteration $(\kappa + 1)$ is given by
\[
\min_{P_T,\omega,\delta, T_{cm}}, \sum_{k \in K} P_{T,k}, \quad (16a)
\]
\[
\text{s.t.} \quad \sum_{k \in K} \nu_{\kappa} M_{ck} \omega_k \left(\omega_k + \tilde{\omega}_k\right)^2 \leq \delta \eta P_{BS_i} \left|q_k(\psi^{(\kappa)})\right|^2 \tau, \quad \forall k \in K, \quad (16b)
\]
\[
(8f), (8g), (10), (12a), (13), (14), (15a), (15b), \quad (16c)
\]
\[
0 < \delta, \delta < \mu, 0 \leq P_{T,k}, 0 \leq T_{cm,k}, \forall k \in K. \quad (16d)
\]

### B. Aggregation Beamforming Optimization

Since (8h) is a non-convex constraint, it is relaxed as $|v|^2 \leq 1$.

At iteration $(\kappa + 1)$, we solve the following approximate convex problem for given $(P_T^{(\kappa+1)}, T_{cm}^{(\kappa+1)}, \psi^{(\kappa)})$:
\[
\max_{\psi} \left(2\Re \left\{ \left(\psi^{(\kappa)}\right)^* \psi \right\} - |\psi^{(\kappa)}|^2 \right) - 1 \quad (17a)
\]
\[
\left(2\Re \left\{ \left(\psi^{(\kappa)}\right)^* Y_k \right\} \Re \left\{ v^* Y_k \right\} - \left|\left(\psi^{(\kappa)}\right)^* Y_k \right|^2 \right) P_{T,k}^{(\kappa+1)}
\]
\[
\text{s.t.} \quad \sum_{\kappa \in K} \frac{\mu_{\kappa} M_{ck}}{2} \omega_k \left(\omega_k + \tilde{\omega}_k\right)^2 \leq \delta^{(\kappa+1)} \eta P_{BS_i} \left|q_k^{(\kappa)}(\psi)\right|^2 \tau, \quad \forall k \in K, \quad (17b)
\]
\[
|v|^2 \leq 1. \quad (17c)
\]

### C. Phase Shift Optimization

Since (8i) is a non-convex constraint, it is relaxed as $|\psi_n|^2 \leq 1, \forall n \in N$.

At iteration $(\kappa + 1)$, we solve the following approximate convex problem for given $(P_T^{(\kappa+1)}, \psi^{(\kappa)}), T_{cm}^{(\kappa+1)}, \omega^{(\kappa+1)}, \delta^{(\kappa+1)}, \mu^{(\kappa+1)})$:
\[
\max_{\psi} \sum_{n \in N} \left(2\Re \left\{ \left(\psi^{(\kappa)}\right)^* \psi_n \right\} - |\psi_n^{(\kappa)}|^2 \right) - N \quad (18a)
\]
\[
\frac{\nu_{\kappa} M_{ck} \omega_k \left(\omega_k + \tilde{\omega}_k\right)^2}{2} \leq \delta^{(\kappa+1)} \eta P_{BS_i} \left|q_k^{(\kappa)}(\psi)\right|^2 \tau, \quad \forall k \in K, \quad (18b)
\]
\[
P_{T,k}^{(\kappa+1)} T_{cm,k} \leq \mu^{(\kappa+1)} \eta P_{BS_i} \left|q_k^{(\kappa)}(\psi)\right|^2 \tau, \quad \forall k \in K, \quad (18c)
\]
\[
|\psi_n|^2 \leq 1, \forall n \in N, \quad (18d)
\]

where
\[
\varphi_k^{(\kappa)}(\psi) = \left(2\Re \left\{ q_k^{(\kappa)}(\psi) \right\} q_k(\psi) \right) - \left|q_k(\psi^{(\kappa)})\right|^2.
\]

The proposed alternating-based iterative algorithm for solving problem (8) is summarized in Algorithm 1.

**Algorithm 1 Proposed Alternating-Based Iterative Algorithm to Solve Problem (8)**

**Initialization:** Set $\kappa := 0$ and generate an initial feasible point $(P_T^{(0)}, \psi^{(0)}, T_{cm}^{(0)}, \psi^{(0)})$.

**repeat**

Given $(P_T^{(\kappa)}, \psi^{(\kappa)}, T_{cm}^{(\kappa)}, \psi^{(\kappa)})$, solve the convex problem (16) to find the set of optimal solutions $(P_T^{*}, \omega^{*}, \delta^{*}, T_{cm}^{*}, \mu^{*})$ and update $P_T^{(\kappa+1)} := P_T*, \omega^{(\kappa+1)} := \omega^*, \delta^{(\kappa+1)} := \delta^*, T_{cm}^{(\kappa+1)} := T_{cm}^*$.

Given $(P_T^{(\kappa+1)}, T_{cm}^{(\kappa+1)}, \psi^{(\kappa)})$, solve the convex problem (17) to find the optimal solution $\psi^*$ and update $\psi^{(\kappa+1)} := \psi^*$.

Given $(P_T^{(\kappa+1)}, \psi^{(\kappa)}, T_{cm}^{(\kappa+1)}, \omega^{(\kappa+1)}, \delta^{(\kappa+1)}, \mu^{(\kappa+1)})$, solve the convex problem (18) to find the optimal solution $\psi^*$ and update $\psi^{(\kappa+1)} := \psi^*$.

Set $\kappa := \kappa + 1$.

**until** Convergence

**Outpt:** $(P_T^*, \omega^*, \delta^*, \psi^*, T_{cm}^*, \mu^*, \psi^*)$

The per-iteration computational complexity of solving (16), (17), and (18) is $O((6K + 2)^2(8K + 1)^{2.5} + (8K + 1)^{3.5})$, $O((K + 1)^{3.5})$, and $O((2K + N)^{3.5})$, respectively.

### V. Numerical Results

We consider a 2-D scenario, in which two BSs and one RIS are located at $(0, 20)$, $(50, 20)$, and $(20, 0)$, respectively. The total number of MUs $K$ is 8, which are randomly distributed within a circular disc with the center (30, 10) and a radius of 1 m. Without loss of generality, we assume that the system bandwidth $B$ is 20 MHz, the transmit power of BS1 is 10 W, the energy harvesting efficiency $\eta$ is 0.9, and the number of RIS REs $N$ is 64. For local computation, the maximum and minimum CPU frequencies $\omega_{\max}$ and $\omega_{\min}$ are 1.5 GHz and 0.3 GHz, respectively, for all users [11]. The required data size $\Theta_k$ is set to 30 kbits, the coefficient of computing chip $\nu_k$ is $2 \times 10^{-28}$, and the local data $S$ and the number of CPU cycles $c_k$ are assumed to be 1000 bits and $10^4$, respectively. The maximum number of local iterations at the $k$-th MU $M_k$ is 4, and the time frame $\tau$ is set to 8 s. All simulation results are generated by averaging over 1000 different random channel realizations.

Fig. 2(a) shows the convergence behavior of Algorithm 1. It is clear that Algorithm 1 needs around 5 iterations to reach the optimal value of the total transmit power of participating MUs. Besides, increasing $\Theta_k$ requires more total transmit power of participating MUs.

In Fig. 2(b), this power is illustrated versus the number of MUs, for different phase selection schemes, i.e., optimized, discrete, and random PS. Not surprisingly, the total transmit power of the participating MUs increases with the number of MUs. Additionally, optimized continuous PS achieves a
b) Total transmit power of participating MUs versus the number of MUs.

(a) Convergence of Algorithm 1 with one random channel realization.

(b) With different values of $\Theta_k$.

(c) With different numbers of MUs.

Further, increased total transmit power of participating MUs is needed for higher $\Theta_k$.

VI. CONCLUSION

In this article, we have considered the total transmit power minimization problem of FL-based wireless networks with the assistance of the RIS. It involves a joint optimization of the transmission time, power control, and the RIS’s phase shifts, and is formulated as a non-convex problem. To solve this problem, we have developed an alternating descent algorithm based on the IA framework, which converges at least to a locally optimal solution. Numerical results have verified the quick convergence of the proposed algorithm and the benefit of using RIS.

REFERENCES

[1] F. Musumeci et al., “An overview on application of machine learning techniques in optical networks,” IEEE Commun. Surveys Tuts., vol. 21, no. 2, pp. 1383–1408, 2nd Quart., 2019.

[2] S. Naser et al., “Towards federated learning-enabled visible light communication in 6G systems,” Oct. 2021, arXiv:2110.03319.

[3] W. Y. B. Lim et al., “Federated learning in mobile edge networks: A comprehensive survey,” IEEE Commun. Surveys Tuts., vol. 22, no. 3, pp. 2031–2063, 3rd Quart., 2020.

[4] X. Wang, Y. Chen, and O. A. Dobre, “Federated learning for anomaly detection: A case of real-world energy storage deployment,” in Proc. IEEE Int. Conf. Commun. (ICC), May 2022, pp. 1–6.

[5] H. Kim, J. Park, M. Bennis, and S.-L. Kim, “Blockchain-enabled on-device federated learning,” IEEE Commun. Lett., vol. 24, no. 6, pp. 1279–1283, Jun. 2020.

[6] Y. Lu, X. Huang, Y. Dai, S. Maharjan, and Y. Zhang, “Blockchain and federated learning for privacy-preserved data sharing in 5G IoT,” IEEE Trans. Ind. Informat., vol. 16, no. 6, pp. 4177–4186, Jun. 2020.

[7] Y. Liu, J. Nie, X. Li, S. H. Ahmed, W. Y. B. Lim, and C. Miao, “Federated learning in the sky: Aerial-ground air quality sensing framework with UAV swarms,” IEEE Internet Things J., vol. 8, no. 12, pp. 9827–9837, Jun. 2021.

[8] K. Li, H. Zhou, Z. Tu, W. Wang, and H. Zhang, “Distributed network intrusion detection system in satellite-terrestrial integrated networks using federated learning,” IEEE Access, vol. 8, pp. 214852–214865, 2020.

[9] S. Zarandi and H. Tabassum, “Federated double deep Q-learning for joint delay and energy minimization in IoT networks,” in Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops), Jun. 2021, pp. 1–6.

[10] X. Mo and J. Xu, “Energy-efficient federated edge learning with joint communication and computation design,” J. Commun. Inf. Netw., vol. 6, no. 2, pp. 110–124, Jun. 2021.

[11] H.-V. Tran, G. Kaddoum, H. Elgala, C. Abou-Rjeily, and H. Kaushal, “Lightwave power transfer for federated learning-based wireless networks,” IEEE Commun. Lett., vol. 24, no. 7, pp. 1472–1476, Jul. 2020.

[12] Q. Zeng, Y. Du, and K. Huang, “Wirelessly powered federated edge learning: Optimal tradeoffs between convergence and power transfer,” IEEE Trans. Wireless Commun., vol. 21, no. 1, pp. 680–695, Jan. 2022.

[13] Z. Wang, Y. Shi, Y. Zhou, H. Zhou, and N. Zhang, “Wireless-powered over-the-air computation in intelligent reflecting surface aided IoT networks,” IEEE Internet Things J., vol. 8, no. 3, pp. 1585–1598, Feb. 2021.

[14] X. Li, G. Zhu, Y. Gong, and K. Huang, “Wirelessly powered data aggregation for IoT via over-the-air function computation: Beamforming and power control,” IEEE Trans. Wireless Commun., vol. 18, no. 7, pp. 3437–3452, Jul. 2019.

[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] 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.

[17] N. H. Tran, W. Bao, A. Zomaya, M. N. H. Nguyen, and C. S. Hong, “Federated learning over wireless networks: Optimization model design and analysis,” in Proc. IEEE Conf. Comput. Commun. (INFOCOM), Apr./May 2019, pp. 1387–1395.

[18] R. Jin, X. He, and H. Dai, “Communication efficient federated learning with energy awareness over wireless networks,” 2020, arXiv:2004.07351.