2025 International Conference on Intelligent Computing
July 26-29, Ningbo, China
https://www.ic-icc.cn/2025/index.php

Energy Efficiency Maximization in Wireless Federated Learning under Inter-Channel Interference

Xinjie Yuan, Shengjie Zhao (✉), Weichao Chen(✉), Fengxia Han, Jin Zeng and Enze Cui

Engineering Research Center of Key Software Technologies for Smart City Perception and Planning, Ministry of Education, School of Computer Science and Technology, Tongji University, Shanghai 201804, China
shengjiezhao@tongji.edu.cn, carlyle.chen@tongji.edu.cn

Abstract. Federated Learning (FL) offers a collaborative learning paradigm for a large number of devices while avoiding data centralization, which is particularly advantageous in wireless environments. However, inter-channel interference, a factor not fully explored in existing FL studies, significantly impacts model transmission and poses substantial challenges for resource allocation in the Wireless FL (WFL) framework. Additionally, the limited energy budgets of mobile devices necessitate energy-efficient strategies across both local computation and model transmission phases. To address these challenges, we formulate a joint learning and communication optimization problem aimed at maximizing the system's Energy Efficiency (EE) under given constraints. We address the problem by decomposing it into two sub-problems: power allocation and client selection, then tackling them sequentially. First, a designed graph neural network (GNN) is employed to parameterize the power allocation strategy, which is optimized through a primal-dual algorithm. Based on the power allocation model, we propose an online algorithm for energy-efficient client selection. Experimental results demonstrate that the proposed method achieves superior EE and reduced energy consumption compared to three baseline methods, while ensuring high-quality wireless transmission and achieving comparable global model accuracy.

Keywords: Federated Learning, Graph Neural Networks, Client Selection, Power Allocation, Energy Efficiency.

1 Introduction

The Internet of Things (IoT) and mobile devices generate vast amounts of daily data, providing valuable training samples for machine learning (ML) applications [1]. However, collecting and storing this data is costly, and it raises significant privacy and security concerns. Furthermore, due to limited wireless resources, it is impractical for edge devices to transmit all their data to the cloud. FL allows mobile devices to train ML models locally, coordinated by a central server, while keeping all training data on the devices. This approach eliminates the need to upload and store data in the cloud, addressing privacy, security, and resource limitations [2].

In WFL, the base station (BS) typically functions as the central server, while mobile devices act as clients. At the start of each WFL round, the BS selects a subset of devices and transmits the global model to them. The devices then train the model using their local data and upload the updated models back through wireless networks to the BS for aggregation.

Since mobile devices are typically powered by batteries, their limited energy constrains the number of rounds they can participate in FL [3]. Moreover, due to the scarcity of wireless resources, only a limited number of devices can participate in FL during each training round. Therefore, the BS must strategically select a subset of devices to participate in each training round. Most existing client selection algorithms focus on optimizing the performance of the global FL model by selecting clients with more data samples [4] or higher local loss [5]. Additionally, [6] proposes an iterative algorithm to minimize the energy consumption of selected clients in FL. However, the majority of existing studies focus on enhancing FL performance and convergence speed, while overlooking the energy consumption of edge devices.

Since the transmission of FL models relies on wireless networks, network conditions and the allocation of wireless resources significantly impact the final performance of the FL model [7]. Therefore, many studies have begun designing FL frameworks for joint device scheduling and wireless resource allocation [8, 9]. However, existing wireless FL frameworks typically assume transmission over interference-free orthogonal channels [6, 8, 10], which often results in devices being allocated the maximum transmission power. In real wireless communication scenarios, such algorithms may exhibit instability due to the presence of inter-channel interference [11].

In addressing the wireless power allocation problem, GNN-based methods have demonstrated superior performance compared to traditional approaches [12]. GNNs are particularly well-suited for wireless networks due to their ability to efficiently learn from graph-structured data. Existing studies have leveraged GNNs to optimize metrics such as the total capacity of wireless communication [13] or the packet error rate (PER) [11]. However, by neglecting the energy consumption of wireless transmission, these models tend to prioritize transmission performance over energy efficiency. Unlike existing research, this paper considers wireless FL under interference constraints and aims to optimize EE in both the model training and wireless communication phases. We formulate an optimization problem aimed at maximizing the overall EE of FL and then decompose it into two sub-problems: power allocation and client selection. Then, we address these two sub-problems sequentially. The main contributions of this paper are as follows:

- We develop a power allocation algorithm based on GNN to optimize EE during the wireless transmission phase, while ensuring high transmission quality in environments affected by inter-channel interference.
- Utilizing the power allocation model, we design a client selection algorithm under constraints such as training time limits to optimize the EE during the computation phase.

- Through a series of numerical experiments on simulated wireless channels and benchmark machine learning datasets, we validate the efficiency and effectiveness of the proposed WFL algorithm.

2 SYSTEM MODEL AND PROBLEM FORMULATION

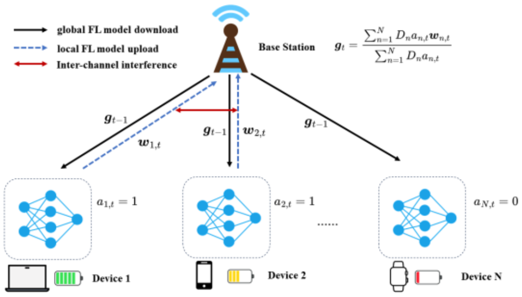

Fig. 1. Wireless Federated Learning for Battery-Powered Mobile Devices.

2.1 Federated Learning over Wireless Networks

As shown in **Fig. 1**, we consider an FL system consisting of a base station (BS) with n_R antennas and N single-antenna devices, where the BS and the devices communicate over a wireless network. Each device has a local dataset $\mathcal{D}_n$ containing D_n data samples. In each dataset $\mathcal{D}_n = \{\boldsymbol{x_n}(i), y_n(i)\}_{i=1}^{D_n}$, $\boldsymbol{x_n}(i) \in R^d$ represents the input vector, and $y_n(i)$ represents the corresponding output.

At the beginning of each FL round, the BS selects N_0 devices $(N_0 \leq N)$ to participate in this round and transmits the global model to them. The selected devices then use their local datasets $\mathcal{D}_n$ to train a parameterized model $\Phi(\cdot; \boldsymbol{w_n})$ to achieve the mapping $\hat{y}_n = \Phi(\boldsymbol{x}_n; \boldsymbol{w}_n)$. Each device's objective is to minimize the loss function $f(\boldsymbol{x}_n, y_n; \boldsymbol{w}_n) \coloneqq \sum_{i=1}^{D_n} \ell(\hat{y}_n(i), y_n(i))$. The loss function ℓ varies depending on the learning task, such as cross-entropy for classification or mean squared error (MSE) for regression. Once the t-th training round is complete, the device n uploads its local model $\boldsymbol{w_{n,t}}$ to the BS via the wireless network and the BS performs model aggregation. The global model update after the t-th round of training is given by

$$\boldsymbol{g_t} = \frac{\sum_{n=1}^{N} D_n a_{n,t} \boldsymbol{w_{n,t}}}{\sum_{n=1}^{N} D_n a_{n,t}}, \tag{1}$$

where $a_{n,t} \in \{0,1\}$ indicates whether device n is selected in the t-th round of training.

Let the CPU frequency of device n at round t be $f_{n,t}$. The computation time for device n is given by

$$\tau_{n,t}^{\text{comp}} = \frac{a_{n,t}\varepsilon_n D_n \omega_n}{f_{n,t}}, \tag{2}$$

where ε_n denotes the number of CPU cycles required by device n to process each bit of sample data and ω_n indicates the number of bits per data sample. According to Lemma 1 in [14], the energy consumption for computation in each round is given by

$$E_{n,t}^{\text{comp}} = a_{n,t}\delta_n \omega_n \varepsilon_n D_n f_{n,t}^2, \tag{3}$$

where δ_n represents the energy coefficient of device n. Each device has an energy budget E_n which represents its total energy available for FL.

2.2 Wireless Transmission with Interference

We focus on the uplink communication process from the devices to the BS. After the t-th round of training, device n transmits its model data to the BS with power $p_{n,t}$. Let $s_n \in \mathbb{C}$ denote the signal sent by device n, and the signal received at the BS is $r = \sum_{n=1}^{N} \mathbf{h}_n s_n + \mathbf{z}$, where $\mathbf{h}_n \in \mathbb{C}^{n_R}$ is the channel from device n to the BS, and $\mathbf{z} \sim \mathcal{N}_{\mathcal{C}}(0, \sigma_n^2)$ is the additive complex Gaussian noise. The signal-to-interference-plus-noise ratio (SINR) of device n to the BS in the t-th round can be defined as

$$\text{SINR}_{n,t} = \frac{\alpha_n p_{n,t}}{1 + \sum_{m \neq n} \beta_{n,m} p_{m,t}}, \tag{4}$$

where $\alpha_n = \parallel \mathbf{h}_n \parallel^2 / \sigma_n^2$ is the channel gain of device n and $\beta_{n,m} = |\mathbf{h}_n^H \mathbf{h}_m|^2 / \sigma_n^2 \parallel \mathbf{h}_n \parallel^2$ is the interference coefficient between devices n and m. Then, we can define the channel state information (CSI) matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, where $\mathbf{H}_{n,n} = \alpha_n$ and $\mathbf{H}_{n,m} = \beta_{n,m}$ when $n \neq m$.

Let $\mathbf{p}_t = (p_{1,t}, \dots, p_{n,t}) \in \mathbb{R}^n$ be the power allocation vector at round t, and $\mathbf{H} \in \mathcal{H} \subseteq \mathbb{R}_+^{n \times n}$ be the CSI matrix drawn from a specific channel distribution $m(\mathbf{H})$. We can calculate the transmission time $\tau_{n,t}^{\text{trs}}$ of device n as

$$\tau_{n,t}^{\text{trs}}(\mathbf{p}_t, \mathbf{H}) = \frac{M}{B \log(1 + \text{SINR}_{n,t})}, \tag{5}$$

where B is the bandwidth, and M is the size of the model transmitted by device n (in bits). The energy consumption during the wireless transmission phase is given by

$$E_{n,t}^{\text{trs}} = p_{n,t}\tau_{n,t}^{\text{trs}}. \tag{6}$$

For simplicity, we assume that each model is transmitted in a single packet [8]. Then the packet error rate (PER) is defined as

$$\text{PER}_{n,t}(\mathbf{p}_t, \mathbf{H}) = (1 - e^{-\frac{c}{\text{SINR}_{n,t}}}), \tag{7}$$

where c is the waterfall threshold [15]. If the transmission of model $\boldsymbol{w}_{n,t}$ fails, it will not be aggregated by the BS. Therefore, the model aggregation equation (1) can be updated as

$$\boldsymbol{g}_t = \frac{\sum_{n=1}^{N} D_n a_{n,t} S_{n,t}(\mathbf{p}_t, \mathbf{H}) \boldsymbol{w}_{n,t}}{\sum_{n=1}^{N} D_n a_{n,t} S_{n,t}(\mathbf{p}_t, \mathbf{H})} \tag{8}$$

where

$$S_{n,t}(\mathbf{p}, \mathbf{H}) = \begin{cases} 1, & \text{with probability } 1 - \mathrm{PER}_{n,t}(\mathbf{p}_t, \mathbf{H}), \\ 0, & \text{with probability } \mathrm{PER}_{n,t}(\mathbf{p}_t, \mathbf{H}). \end{cases} \tag{9}$$

2.3 Problem Formulation

When considering the contribution of a device to FL, two main factors are primarily taken into account. First, a larger local dataset size indicates a higher likelihood that the device provides high-quality samples [16]. Additionally, a larger local loss value during the training process suggests that the current local model requires more training in subsequent rounds [4]. Based on these two considerations, we define the contribution of device n to FL in round t as

$$v_{n,t} = \sqrt{D_n} f(\boldsymbol{x}_n, y_n; \boldsymbol{w}_{n,t}). \tag{10}$$

The square root operation applied to D_n is intended to mitigate potential unfairness that may arise from significant differences in data sizes. The EE of the computation phase in the t-th FL round is defined as the ratio of the contribution of the selected devices to the energy consumed for computation

$$\mathrm{EE}_t^{\mathrm{comp}} = \frac{\sum_{n=1}^{N} a_{n,t} v_{n,t}}{\sum_{n=1}^{N} a_{n,t} E_{n,t}^{\mathrm{comp}}}. \tag{11}$$

In wireless transmission systems, EE is typically defined as the ratio of successfully transmitted bits to the energy consumed in transmission [17]. Based on this definition, the EE of the wireless transmission phase for device n is expressed as

$$\mathrm{EE}_{n,t}^{\mathrm{trs}} = \frac{a_{n,t} M}{E_{n,t}^{\mathrm{trs}}}. \tag{12}$$

The optimization objective is to maximize the combined energy efficiency of computation and transmission in each round of FL, while adhering to the necessary constraints of both phases. The entire optimization problem is defined as follows

$$
\begin{aligned}
P1 \quad \max_{\mathbf{a}_t,\mathbf{p}_t,\mathbf{f}_t} \quad & \mathrm{EE}_t^{\mathrm{comp}} + \sum_{n=1}^{N} a_{n,t}\mu_n \mathrm{EE}_{n,t}^{\mathrm{trs}}, \forall t \in \mathcal{T} && (13a) \\
\text{s.t.} \quad & a_{n,t} \in \{0,1\}, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}, && (13b) \\
& \sum_{n=1}^{N} a_{n,t} \le N_0, \forall t \in \mathcal{T}, && (13c) \\
& f_{min} \le a_{n,t} f_{n,t} \le f_{max}, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}, && (13d) \\
& a_{n,t}\tau_{n,t}^{\mathrm{comp}} \le \tau_0^{\mathrm{comp}}, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}, && (13e) \\
& p_{\min} \le a_{n,t} p_{n,t} \le p_{\max}, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}, && (13f) \\
& a_{n,t}\tau_{n,t}^{\mathrm{trs}} \le \tau_0^{\mathrm{trs}}, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}, && (13g) \\
& a_{n,t}\mathrm{PER}_{n,t} \le q_0, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}. && (13h) \\
& a_{n,t}(E_{n,t}^{\mathrm{comp}} + E_{n,t}^{\mathrm{trs}}) \le E_0, \forall n \in \mathcal{U}, \forall t \in \mathcal{T}. && (13i)
\end{aligned}
$$

Here, $\mathbf{a}_t = (a_{1,t}, \dots, a_{n,t}), \mathbf{p}_t = (p_{1,t}, \dots, p_{n,t}), \mathbf{f}_t = (f_{1,t}, \dots, f_{n,t})$ represents the client selection vector, the power allocation vector and the CPU frequency vector, respectively. Let $\mathcal{U} = \{1, \dots, \mathrm{N}\}$ and $\mathcal{T} = \{1, \dots, \mathrm{T}\}$ be the sets of device indices and training rounds, respectively. The parameter μ_n is introduced to represent the energy efficiency weight of the device n during the wireless communication phase, which is set to $\frac{1}{N}$ in the experiment. Constraint (13d) specifies the range of CPU frequencies for the selected devices, while Constraint (13e) imposes time limits on the training phase. Constraint (13f) bounds the transmission power between $p_{\min}$ and $p_{\max}$. The constraints in (13g) and (13h) ensure that for each selected device, the expected transmission time and packet error rate remain within the given thresholds. The constraint in (13i) ensures that the energy consumption of the selected device in each round does not exceed the energy budget E_0.

3 Proposed Method

Addressing problem $P1$ directly poses considerable challenges due to several factors: the non-convex nature of both the objective function and the constraints, the channel state information which follows a specific distribution and dynamically changes in wireless transmission, and the interdependence of multiple constraints. By recognizing the monotonicity of the objective function with respect to f, we first derive a closed-form solution for the frequency allocation problem. This closed-form solution enables us to decouple the original problem into two subproblems—power allocation and client selection—which are then solved sequentially.

3.1 Power allocation

We observe that the objective function is monotonically decreasing with respect to the device CPU frequency $\mathbf{f}_t$, while the constraint (13e) essentially imposes a lower bound on the frequency of each device, i.e.,

$$a_{n,t}f_{n,t} \geq \frac{\varepsilon_n D_n \omega_n}{\tau_0^{comp}}. \tag{14}$$

Therefore, we can derive a closed-form expression for determining the optimal CPU frequencies of the selected devices, which is presented as follows

$$f_n^* = \min\left(\max\left(f_{\min}, \frac{\varepsilon_n D_n \omega_n}{\tau_0^{\text{comp}}}\right), f_{\max}\right) \tag{15}$$

By substituting (15), we can rewrite constraint (13i) in a form that only involves $E_{n,t}^{\text{trs}}$. Then, we derive the power allocation subproblem by adopting a fixed selection strategy. Specifically, given the client selection policy a_t, we can derive the CSI matrix $\mathbf{H}'$ of the selected devices from H. Subsequently, the problem is reformulated as a power allocation subproblem based on $\mathbf{H}'$. The objective of the power allocation subproblem is to determine the power function for each instantaneous channel state, i.e., $p(\mathbf{H}')$, in order to optimize the devices' EE while satisfying the constraints in wireless transmission. Let $e_0 = E_0 - \delta_n \omega_n \varepsilon_n D_n (f_n^*)^2$, the power allocation subproblem $S1$ is written as

$$\begin{aligned}
S1 \quad \max_{\mathbf{p}} \quad & \sum_{n=1}^{N_0} \mu_n \mathbb{E}_{\mathbf{H}'}[\text{EE}_n^{\text{trs}}(\mathbf{p}, \mathbf{H}')], && \text{(16a)}\\
\text{s.t.} \quad & p_{\min} \leq p_n(\mathbf{H}') \leq p_{\max}, \forall n \in \mathcal{U}, && \text{(16b)}\\
& \mathbb{E}_{\mathbf{H}'}[\tau_n^{\text{trs}}(\mathbf{p}, \mathbf{H}')] \leq \tau_0^{\text{trs}}, \forall n \in \mathcal{U}, && \text{(16c)}\\
& \mathbb{E}_{\mathbf{H}'}[\text{PER}_n(\mathbf{p}, \mathbf{H}')] \leq q_0, \forall n \in \mathcal{U}, && \text{(16d)}\\
& \mathbb{E}_{\mathbf{H}'}[E_n^{\text{trs}}(\mathbf{p}, \mathbf{H}')] \leq e_0, \forall n \in \mathcal{U}. && \text{(16e)}
\end{aligned}$$

It is worth mentioning that since $\mathbf{H}$ is drawn from a certain channel distribution $\mathcal{H}$, the function related to $\mathbf{H}'$ in $S1$ is expressed as expectation. This formulation aims to optimize the power allocation strategy $\mathbf{p}$ under dynamically varying channel conditions.

Due to the non-convex nature of the problem and the infinite-dimensionality of the power allocation function $\mathbf{p}(\mathbf{H}')$, directly solving $S1$ is highly challenging. Inspired by [13], we employ a specifically designed GNN to parameterize the power allocation function, reducing the dimension of sub-problem $S1$. Specifically, we define

$$\mathbf{p}(\mathbf{H}') = \phi(\mathbf{H}'; \boldsymbol{\theta}), \tag{17}$$

where the CSI matrix $\mathbf{H}'$ is treated as the adjacency matrix of the sub-graph corresponding to the selected device. ϕ is our designed power allocation graph neural network (PAGNN) and $\boldsymbol{\theta}$ represents the corresponding trainable weights. The input of PAGNN is $\mathbf{F}^{(0)} = p_{\min}\mathbf{1}$ with the output being the power allocation vector $\mathbf{p}$.

The first part of the PAGNN employs graph attention network (GAT) to capture non-local dependencies and importance weights among fading channels. GAT uses a self-attention mechanism to assign different attention weights to each node, thereby enabling better handling of complex interactions between nodes [18]. After GAT, we employ multiple layers of graph convolutional network (GCN). GCNs learn the power allocation function by aggregating information from neighboring nodes and performing feature extraction through convolution operations. The l-th intermediate layer of the GCN is defined as

$$\mathbf{F}^{(l+1)} = \sigma\left(\mathbf{D}^{-\frac{1}{2}}\widetilde{\mathbf{H}}\mathbf{D}^{-\frac{1}{2}}\mathbf{F}^{(l)}\mathbf{W}^{(l)}\right), \tag{18}$$

where $\widetilde{\mathbf{H}} = \mathbf{H}' + I$ with I being the identity matrix, $\mathbf{D}$ is the degree matrix of $\widetilde{\mathbf{H}}$, $\mathbf{W}^{(l)}$ is the weight matrix of the l-th layer, and σ is the activation function.

In fact, various neural network architectures can be used as the parameterization method for power allocation. A key reason for selecting GNNs is their ability to handle inputs of varying sizes, which allows for flexible application in the subsequent client selection sub-problem. Furthermore, GNNs can ensure permutation invariance [19], which guarantees that the power allocation strategy remains stable and unaffected by changes in node indexing.

After parameterizing the power allocation $\mathbf{p}$ with (17), $S1$ is transformed into an optimization problem with respect to $\boldsymbol{\theta}$. We aim to solve $S1$ with standard gradient descent-based algorithms. To achieve this, we first attempt to formulate the Lagrangian dual problem of $S1$. The reformulated problem $S1.1$ is as follows

$$\text{S1.1} \quad \max_{\theta,e,\tau,q,\tilde{e}} \quad \sum_{n=1}^{N} \mu_n e_n, \tag{19a}$$

$$\text{s.t.} \quad \phi_n(\mathbf{H}';\boldsymbol{\theta}) \in [p_{min}, p_{max}], \forall n \in \mathcal{U}, \tag{19b}$$

$$\boldsymbol{e}_n \leq \mathbb{E}_{\mathrm{H}}[\mathrm{EE}_n^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')], \forall n \in \mathcal{U}, \tag{19c}$$

$$\boldsymbol{\tau}_n \geq \mathbb{E}_{\mathrm{H}}[\tau_n^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')], \forall n \in \mathcal{U}, \tag{19d}$$

$$\boldsymbol{q}_n \geq \mathbb{E}_{\mathrm{H}}[\mathrm{PER}_n(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')], \forall n \in \mathcal{U}, \tag{19e}$$

$$\tilde{\boldsymbol{e}}_n \geq \mathbb{E}_{\mathrm{H}}[E_n^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')], \forall n \in \mathcal{U}, \tag{19f}$$

$$\boldsymbol{\tau}_n \in (0, \tau_0^{\mathrm{trs}}], \forall n \in \mathcal{U}, \tag{19g}$$

$$\boldsymbol{q}_n \in [0, q_0], \forall n \in \mathcal{U}, \tag{19h}$$

$$\tilde{\boldsymbol{e}}_n \in (0, e_0], \forall n \in \mathcal{U}. \tag{19i}$$

For the power constraints in (19b), we directly impose the limits within the output of the PAGNN. We then introduce the dual variables $\boldsymbol{e}, \boldsymbol{\tau}, \boldsymbol{q}$ and $\tilde{\boldsymbol{e}}_n$, along with the non-negative Lagrange multipliers $\lambda_{\{e,\tau,q,\tilde{e}\}} \in R_+^N$. For simplicity, we denote $\mathbb{E}_{\mathrm{H}}[\mathrm{EE}^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')]$, $\mathbb{E}_{\mathrm{H}}[\mathrm{PER}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')]$, $\mathbb{E}_{\mathrm{H}}[\boldsymbol{\tau}^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')]$ and $\mathbb{E}_{\mathrm{H}}[E^{\mathrm{trs}}(\phi(\mathbf{H}';\boldsymbol{\theta}), \mathbf{H}')]$ as $\mathbb{E}_{\mathrm{H}}[f_e]$, $\mathbb{E}_{\mathrm{H}}[f_q]$, $\mathbb{E}_{\mathrm{H}}[f_\tau]$ and $\mathbb{E}_{\mathrm{H}}[f_{\tilde{e}}]$ respectively. The Lagrangian function corresponding to $S1.1$ is then defined as follows

$$\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{e}, \boldsymbol{q}, \boldsymbol{\tau}, \tilde{\boldsymbol{e}}, \boldsymbol{\lambda}_e, \boldsymbol{\lambda}_q, \boldsymbol{\lambda}_\tau, \boldsymbol{\lambda}_{\tilde{e}}) = \sum_{n=1}^{N} \mu_n \boldsymbol{e}_n + \boldsymbol{\lambda}_e^{\mathrm{T}}(-\boldsymbol{e} + \mathbb{E}_{\mathrm{H}}[f_e])$$

$$+\boldsymbol{\lambda}_q^{\mathrm{T}}(\boldsymbol{q} - \mathbb{E}_{\mathrm{H}}[f_q]) + \boldsymbol{\lambda}_\tau^{\mathrm{T}}(\boldsymbol{\tau} - \mathbb{E}_{\mathrm{H}}[f_\tau]) + \boldsymbol{\lambda}_{\tilde{e}}^{\mathrm{T}}(\tilde{\boldsymbol{e}} - \mathbb{E}_{\mathrm{H}}[f_{\tilde{e}}]). \tag{20}$$

We also denote the feasible domains of $\boldsymbol{\tau}$, $\boldsymbol{q}$ and $\tilde{\boldsymbol{e}}$ as $\mathcal{X} \coloneqq (0, \tau_0 \mathbf{1}]$, $\mathcal{Y} \coloneqq [0, q_0 \mathbf{1}]$ and $\mathcal{Z} \coloneqq [0, e_0 \mathbf{1}]$. We employ the Lagrangian function as the designated loss function for PAGNN, and, based on the standard dual learning framework [20], we derive the subsequent iterative formula:

$$\begin{aligned}
\boldsymbol{\theta}_{k+1} &= \boldsymbol{\theta}_k + \rho_{\boldsymbol{\theta}}(\nabla_\theta \mathbb{E}_{\mathrm{H}}[f_e]\boldsymbol{\lambda}_{e,k} - \nabla_\theta \mathbb{E}_{\mathrm{H}}[f_q]\boldsymbol{\lambda}_{q,k} \\
&\quad - \nabla_\theta \mathbb{E}_{\mathrm{H}}[f_\tau]\boldsymbol{\lambda}_{\tau,k} - \nabla_\theta \mathbb{E}_{\mathrm{H}}[f_{\tilde{e}}]\boldsymbol{\lambda}_{\tilde{e},k}), && (21a) \\
\boldsymbol{e}_{k+1} &= \boldsymbol{e}_k + \rho_{\boldsymbol{e}}(\boldsymbol{\mu} - \boldsymbol{\lambda}_{e,k}), && (21b) \\
\boldsymbol{q}_{k+1} &= \mathrm{proj}_{\mathcal{X}}\,(\boldsymbol{q}_k + \rho_{\boldsymbol{q}}\boldsymbol{\lambda}_{q,k}), && (21c) \\
\boldsymbol{\tau}_{k+1} &= \mathrm{proj}_{\mathcal{Y}}(\boldsymbol{\tau}_k + \rho_{\boldsymbol{\tau}}\boldsymbol{\lambda}_{\tau,k}), && (21d) \\
\tilde{\boldsymbol{e}}_{k+1} &= \mathrm{proj}_{\mathcal{Z}}\,(\tilde{\boldsymbol{e}}_k + \rho_{\tilde{\boldsymbol{e}}}\boldsymbol{\lambda}_{\tilde{e},k}), && (21e) \\
\boldsymbol{\lambda}_{e,k+1} &= [\boldsymbol{\lambda}_{e,k} - \rho_{\lambda_e}(-\boldsymbol{e} + \mathbb{E}_{\mathrm{H}}[f_e])]_+, && (21f) \\
\boldsymbol{\lambda}_{q,k+1} &= [\boldsymbol{\lambda}_{q,k} - \rho_{\lambda_q}(\boldsymbol{q} - \mathbb{E}_{\mathrm{H}}[f_q])]_+, && (21g) \\
\boldsymbol{\lambda}_{\tau,k+1} &= [\boldsymbol{\lambda}_{\tau,k} - \rho_{\lambda_\tau}(\boldsymbol{\tau} - \mathbb{E}_{\mathrm{H}}[f_\tau])]_+, && (21h) \\
\boldsymbol{\lambda}_{\tilde{e},k+1} &= [\boldsymbol{\lambda}_{\tilde{e},k} - \rho_{\lambda_{\tilde{e}}}(\tilde{\boldsymbol{e}} - \mathbb{E}_{\mathrm{H}}[f_{\tilde{e}}])]_+. && (21i)
\end{aligned}$$

Here we introduce $\rho_{\boldsymbol{\theta},\boldsymbol{e},\boldsymbol{q},\boldsymbol{\tau},\tilde{\boldsymbol{e}},\lambda_e,\lambda_q,\lambda_\tau,\lambda_{\tilde{e}}} > 0$ as scalar step sizes. The gradient updates in (21a) - (21i) constitute a primal-dual learning approach that trains the PAGNN model to maximize wireless transmission energy efficiency while ensuring the satisfaction of constraints.

3.2 Client Selection

After completing the training process of PAGNN, we use it to assist in solving the original problem $P1$. However, using PAGNN to predict power allocation for every client selection strategy a_t incurs significant computational costs. To address this, we propose a more efficient approach. Initially, we set $a_t = 1$ and utilize PAGNN to pre-allocate power $\mathbf{p}$ for all devices. Based on $\mathbf{p}$, we then estimate each device's relevant metrics during the wireless transmission phase, and subsequently solve the client selection subproblem

$$\begin{aligned}
S2 \quad \max_{\mathbf{a}_t} \quad & \frac{\sum_{n=1}^{N} a_{n,t} v_{n,t}}{\sum_{n=1}^{N} a_{n,t} E_{n,t}^{\mathrm{comp}}} + \sum_{n=1}^{N} a_{n,t}\mu_n \mathrm{EE}_{n,t}^{\mathrm{trs}}, && (22a) \\
s.t. \quad & (13b), (13c), (13f) - (13i). && (22b)
\end{aligned}$$

One method for obtaining $v_{n,t}$ is forcing all devices to perform a local training epoch at the start of each FL round and upload the most recent loss value to the BS. However, this introduces additional computational and communication costs. To address this, we compute $v_{n,t}$ using the historical loss from the device's most recent training round, and introduce the Age of Update (AoU) metric to reflect the timeliness of model updates [21]. We define the client n's AoU at round t as AoU_n^t, which can be calculated by

$$\mathrm{AoU}_n^t = \begin{cases} \mathrm{AoU}_n^{t-1} + 1, a_{n,t-1} = 0, \\ 1, a_{n,t-1} = 1. \end{cases} \tag{23}$$

The AoU metric provides an intuitive indication of the device's obsolescence. Based on the theoretical derivation in [22], maximizing the AoU-weighted contribution during the client selection process minimizes the gap between the global model's loss and its expected upper bound. In other words, selecting devices with higher AoU values in each round enables more information updates, which facilitates the convergence of FL. Therefore, we update $v_{n,t}$ in an AoU-weighted manner as follows

$$v_{n,t} = \sqrt{D_n} f(\boldsymbol{x}_n, y_n; \boldsymbol{w}_{\tilde{n},t}) \cdot \frac{\mathrm{AoU}_n^t}{\sum_{i=1}^{N} \mathrm{AoU}_i^t}, \tag{24}$$

where $\boldsymbol{w}_{\tilde{n},t}$ represents the model from the device's most recent local training.

Algorithm 1. Energy-Efficient Client Selection and Power Allocation Algorithm

Input: Device data size vector $\mathcal{D}$, set of devices $\mathcal{U}$, current CSI matrix $\boldsymbol{H}$
Output: Selected device vector $\boldsymbol{a}$, power allocation vector $\mathbf{p}'$

```
p ← φ(H; θ)
Calculate τ^trs_{n,t}, τ^comp_{n,t}, E^trs_{n,t} and E^comp_{n,t} for each device based on (2), (3), (5), (6)
U_sub ← Filter(U)
Update v_{n,t} based on (24)
if |U_sub| ≥ N_0 then
    a ← B&B(U_sub, D, E)
else
    a ← Select(U_sub)
end if
H' ← Transform(H, a)
p' ← φ(H'; θ)
```

$S2$ is a binary nonlinear integer programming problem that does not possess low complexity solutions in general. For cases with a relatively small N, we use the standard branch-and-bound ($B\&B$) algorithm to solve it, while employing strategies such as bound pruning to improve solving efficiency. The original problem $P1$ is solved with the following steps. At the beginning of each FL round, the BS collects the current channel state information H and attempts to allocate power to all devices by PAGNN, i.e., $\mathbf{p} = \phi(\mathbf{H}; \boldsymbol{\theta})$. The BS utilize $\mathbf{p}$ to estimate the relevant performance metrics of devices during the wireless transmission phase and apply the constraint (13e) – (13i) to filter suitable devices. Furthermore, devices with $p_n < p_{\min}$ are also excluded, as this indicates that the model perceives the wireless channel of the device to be poor or that it exhibits strong interference with other channels. The binary integer programming problem is then solved with the $B\&B$ method on the remaining candidate devices. Finally, based on the client selection strategy a_t, the submatrix $\mathbf{H}'$ is derived, and the final device power allocation strategy $\mathbf{p}'$ is obtained using PAGNN. The comprehensive algorithm is outlined in Algorithm 1.

During each round of FL, the BS employs **Algorithm 1** to efficiently schedule devices and allocate resources, subsequently aggregating the local models uploaded by these devices. Upon the completion of T training rounds, the BS provides the global model $\boldsymbol{w}_T$ as the comprehensive final output.

3.3 Convergence Analysis

In this section, we present a convergence analysis of the proposed algorithm. Denote the optimum of original problem $S1$ and dual problem $S1.1$ as P^* and D_ϕ^*, respectively. Define ζ as the upper bound on the approximation error when the parameterized function $\phi(\mathbf{H}'; \boldsymbol{\theta})$ approximates the power allocation function $\mathbf{p}(\mathbf{H}')$. Based on the theoretical proof in [23], under the necessary conditions (such as $\mathcal{H}$ being a nonatomic distribution and the feasible region of the objective function containing at least one interior point), the duality gap $P^* - D_\phi^*$ has an upper bound, i.e.,

$$P^* - D_\phi^* \leq |\lambda^*|_1 K\zeta, \tag{25}$$

where $\|\lambda^*\|_1 \leq \frac{P^* - \mathbf{x}_0}{s} < \infty$, with x_0 representing an interior point value within the feasible region, and s is the strict parameter that causes the objective function to surpass $\mathbf{x}_0$. This conclusion indicates that the approximation error ζ of the parameterized function ϕ has a linear relationship with the duality gap. In other words, as the approximation error ζ of the parameterized function ϕ gradually decreases, the algorithm will progressively converge and achieve a solution close to the optimal solution of $P1$.

3.4 Complexity Analysis

This subsection presents an analysis of the complexity of the proposed algorithm.
The time complexity for each layer of GAT and GCN is $O(E)$, where E denotes the number of edges in the graph. For smaller device scales, the CSI matrix H is typically dense, hence the time complexity of PAGNN can be considered as $O(LN^2)$, with L representing the number of layers in PAGNN. The $B\&B$ algorithm has a worst time complexity of $O(Cb^d)$, where C is treated as a constant, b represents the branching factor (2 in this study), and d indicates the depth of the search tree. Consequently, the overall complexity of the algorithm is $O(LN^2 + C2^d)$. Despite the exponential nature of $B\&B$'s worst-case time complexity, the integration of cutting planes and pruning techniques greatly enhances the algorithm's efficiency, making it feasible for moderate values of N [24]. For extensive FL systems, we recommend heuristic algorithms as they may provide near-optimal solutions within polynomial time complexity.

4 NUMERICAL EXPERIMENTS

In this section, we present a series of experiments to evaluate the performance of the proposed model and algorithm.

4.1 Simulation Settings

Our simulation system consists of one BS and N mobile devices within a $500\mathrm{m} \times 500\mathrm{m}$ area. We generate fading channels following the path loss model from [25] with carrier frequency of 1.8 GHz and power decay factor equal to 4.5. Meanwhile, fast fading is modeled as a circularly symmetric complex Gaussian random variable with zero mean and unit variance. In the system with N devices, a total of 4,000 channel realizations were generated, with 3,000 for training and 1,000 for testing. The architecture of the proposed PAGNN model is as follows: the first hidden layer is a GAT layer with a feature dimension of 8 and 8 attention heads. It is followed by four GCN layers with feature dimensions $\{32, 64, 16, 4\}$. The activation function σ for the intermediate layers is set as ELU. The key parameter values discussed in the text are enumerated in **Table 2**.

Table 2. SYSTEM PARAMETERS

Parameter	Value	Parameter	Value
N	16	n_R	10
ω	25000 bits	c	0.023
ε	30 cycles/bit	$\rho_{\boldsymbol{\theta}}$	5×10^{-4}
δ	10^{-28}	$\rho_{\{e,q,\tau,\tilde{e},\lambda_e,\lambda_q,\lambda_\tau,\lambda_{\tilde{e}}\}}$	1.5×10^{-4}
τ_0^{comp}	10.0 s	τ_0^{trs}	3.0 s
q_0	0.1	batch size	32
$p_{\min}$	10^{-6}W	$f_{\min}$	0.5 GHz
$p_{\max}$	0.1W	$f_{\max}$	2.0 GHz

We first generate channel test sets with different interference intensities and varying numbers of devices to test the effectiveness of PAGNN. We then perform FL experiments using the proposed WFL method on the CIFAR-10 dataset. In each FL round, a new channel realization is generated using the previously described channel fading model to simulate the dynamic wireless network environment. FL training is conducted with a four-layer convolutional neural network (CNN) as the local model, with each experiment comprising 100 FL rounds. Considering data heterogeneity, we partition the dataset among devices according to a logarithmic distribution of dataset sizes, using a log-normal distribution with a standard deviation of 0.7. Additionally, we employ a Dirichlet distribution with the concentration parameter set to 1.0 to achieve a Non-IID partitioning of the dataset across different data categories.

4.2 Effectiveness of PAGNN

First, we evaluate the performance of the proposed PAGNN on the simulated channel dataset by comparing it against three baselines:

- **Rand**: Random power allocation.
- **Orth**: Optimal power allocation under orthogonal channel conditions [8].
- **PDGNet**: A power allocation model based on GCN, optimized to minimize weighted PER [11].

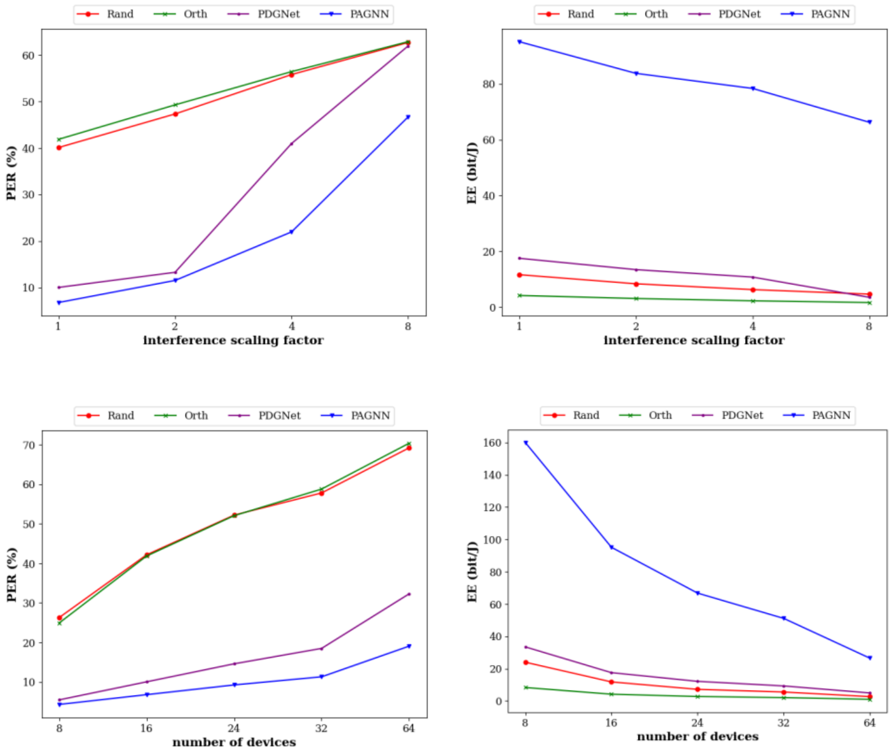

Fig. 2. EE and PER achieved by different power allocation methods under varying interference intensities and network sizes.

We begin by scaling the original wireless network based on the the interference factor, and then compare the EE and PER of devices under varying interference intensities. As shown in **Fig. 2**, both GNN-based methods maintain low error rates under low interference intensity, with PAGNN consistently outperforming other methods. Moreover, under conditions of high interference intensity, PAGNN demonstrates superior stability compared to the other methods. Meanwhile, PAGNN achieves significantly higher energy efficiency than the other methods across various interference intensities, highlighting the effectiveness of the dual learning algorithm.

In experiments with varying numbers of devices, PAGNN consistently achieves the lowest PER and the highest EE, demonstrating its effectiveness for wireless networks of different scales. Although trained solely on channel dataset where $N = 16$, PAGNN effectively adapts to wireless networks of varying scales ($N = \{8,16,24,32,64\}$). This adaptability is due to the inherent flexibility of GNNs in handling inputs of different sizes.

4.3 Comparison of Wireless Federated Learning Methods

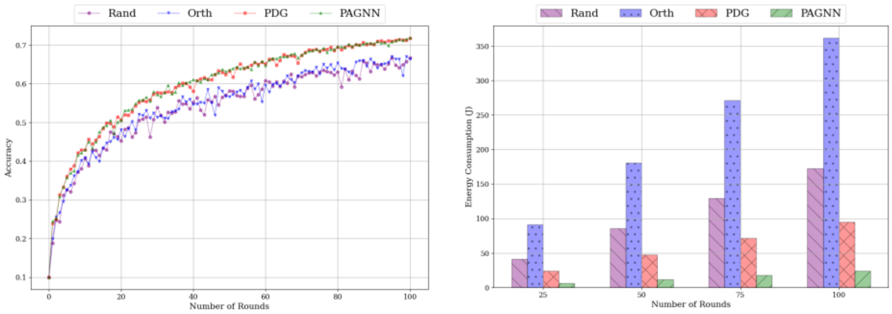

Fig. 3. The performance of different power allocation methods on CIFAR-10 dataset.

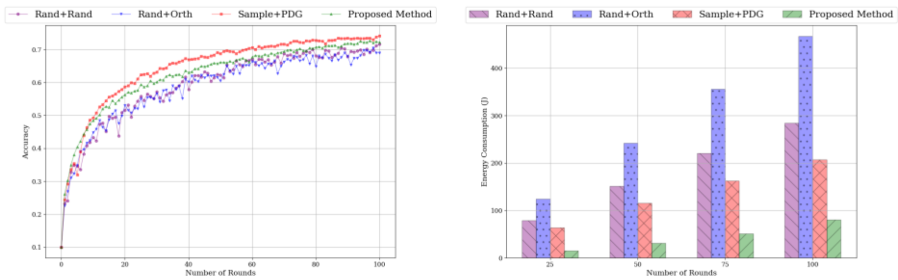

Fig. 4. The performance of different WFL methods on CIFAR-10 dataset with IID data distribution.

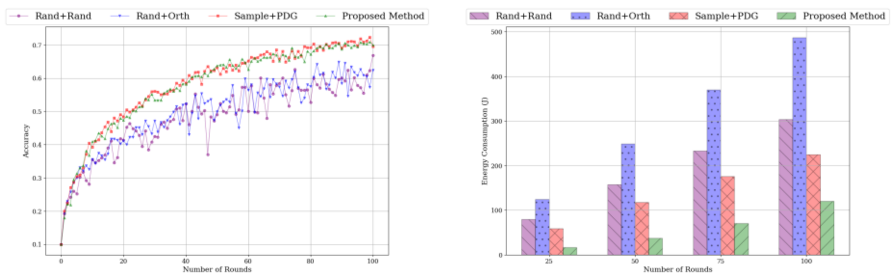

Fig. 5. The performance of different WFL methods on CIFAR-10 dataset with Non-IID data distribution.

Furthermore, to demonstrate the generalization and effectiveness of the proposed FL method, we conducted a series of FL experiments on real-world dataset CIFAR-10. We first conduct FL experiments with variations in the power allocation model only. Next, we perform FL experiments under both IID and Non-IID data distributions using the proposed joint client selection and power allocation algorithm. For experiments with different power allocation model, we fix the client selection strategy to be random selection in FedAvg [26]. For the other two sets of experiments, we establish the following baselines to compare with the proposed algorithm:

- Rand+Rand: Random device selection method used in FedAvg, combined with Rand power allocation strategy. This method serves as a benchmark, simulating conventional FL.
- Rand+Orth: Random device selection with Orth power policy. This method is used to simulate WFL under orthogonal channel conditions.
- Sample+PDG: Probabilistic sampling based on the number of device samples [10], with PDGNet as the power allocation method. This method is set as a high-performance baseline without specific constraints on energy consumption.

The specific experimental results are presented in **Fig. 3**-**Fig. 5**. The results shown in **Fig. 3** illustrate the comparison of FL model accuracy and transmission energy consumption among different power allocation models. It can be observed that FL with PAGNN consistently maintains the lowest transmission energy consumption, while achieving a convergence speed comparable to that of PDGNet. This is attributed to PAGNN's ability to optimize EE while satisfying constraints on transmission time and error rate. Additionally, since the Orth strategy does not account for inter-channel interference, FL with Orth consumes the most energy and exhibits poor performance.

For the IID data distribution, despite the proposed algorithm's convergence speed being slightly slower compared to the Sample+PDG baseline, it ultimately achieves a comparable model accuracy while reducing energy consumption to only 38% of that required by Sample+PDG (**Fig. 4**).

For Non-IID data distribution, as shown in **Fig. 5**, the proposed method demonstrates even better performance, matching the accuracy of the Sample+PDG method while being more stable compared to the other two methods. This is attributed to the introduction of the AoU metric during the client selection phase, which not only ensures the timeliness of model updates but also enables the data from more devices to participate in training. Similar to the case with IID data distribution, the proposed method also achieves the lowest energy consumption under Non-IID distribution, demonstrating its generalizability.

Based on the experimental results above, it can be concluded that the proposed method effectively enhances the transmission quality and EE of WFL in interference environments without making significant compromises in performance.

5 Conclusion

In this paper, we proposed an energy-efficient framework for wireless FL. First, we developed a GNN-based power allocation algorithm to optimize the weighted EE during transmission phase under inter-channel interference. Next, we designed a client selection algorithm aimed at maximizing the EE for computation phase, while ensuring the performance of the FL model. Through numerical experiments on simulated wireless system and benchmark machine learning datasets, we validated that the proposed algorithm enhances EE without compromising FL performance. Future work will extend our model to explore its performance in large-scale IoT scenarios and under more complex communication network conditions.

Acknowledgments. This work was supported in part by the National Key R&D Program of China 2023YFC3806000 and 2023YFC3806002, in part by the National Natural Science Foundation of China under Grant 62401396 and 62076183, in part by Shanghai Municipal Science and Technology Major Project No. 2021SHZDZX0100, in part by the Shanghai Science and Technology Innovation Action Plan Project 22511105300, in part by Natural Science Foundation of Shanghai under Grant 24ZR1467700, in part by the National Natural Science Foundation of China under Grant U23A20382 and in part by Fundamental Research Funds for the Central Universities. The authors would also like to thank the anonymous reviewers for their careful work and valuable suggestions.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Nguyen, D.C., Ding, M., Pathirana, P.N., Seneviratne, A., Li, J., Poor, H.V.: Federated learning for internet of things: A comprehensive survey. IEEE Communications Surveys & Tutorials 23(3), 1622–1658 (2021)
2. Lim, W.Y.B., Luong, N.C., Hoang, D.T., Jiao, Y., Liang, Y.C., Yang, Q., Niyato, D., Miao, C.: Federated learning in mobile edge networks: A comprehensive survey. IEEE communications surveys & tutorials 22(3), 2031–2063 (2020)
3. Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., Jararweh, Y.: Federated learning review: Fundamentals, enabling technologies, and future applications. Information processing & management 59(6), 103061 (2022)
4. Lai, F., Zhu, X., Madhyastha, H.V., Chowdhury, M.: Oort: Efficient federated learning via guided participant selection. In: 15th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 21). pp. 19–35 (2021)
5. Cho, Y.J., Wang, J., Joshi, G.: Towards understanding biased client selection in federated learning. In: International Conference on Artificial Intelligence and Statistics. pp. 10351–10375. PMLR (2022)
6. Yang, Z., Chen, M., Saad, W., Hong, C.S., Shikh-Bahaei, M.: Energy efficient federated learning over wireless communication networks. IEEE Transactions on Wireless Communications 20(3), 1935–1949 (2020)
7. Yang, H.H., Liu, Z., Quek, T.Q., Poor, H.V.: Scheduling policies for federated learning in wireless networks. IEEE transactions on communications 68(1), 317–333 (2019)
8. Chen, M., Yang, Z., Saad, W., Yin, C., Poor, H.V., Cui, S.: A joint learning and communications framework for federated learning over wireless networks. IEEE transactions on wireless communications 20(1), 269–283 (2020)

9. Shi, W., Zhou, S., Niu, Z., Jiang, M., Geng, L.: Joint device scheduling and resource allocation for latency constrained wireless federated learning. IEEE Transactions on Wireless Communications 20(1), 453–467 (2020)
10. Xu, W., Liang, B., Boudreau, G., Sokun, H.: Probabilistic client sampling and power allocation for wireless federated learning. In: 2023 IEEE 34th Annual International Symposium on Personal, Indoor and Mobile Radio Communications(PIMRC). pp. 1–6. IEEE (2023)
11. Li, B., Swami, A., Segarra, S.: Power allocation for wireless federated learning using graph neural networks. In: ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 5243–5247. IEEE (2022)
12. Shen, Y., Shi, Y., Zhang, J., Letaief, K.B.: A graph neural network approach for scalable wireless power control. In: 2019 IEEE Globecom Workshops (GC Wkshps). pp. 1–6 (2019). https://doi.org/10.1109/GCWkshps45667.2019.9024538
13. Eisen, M., Ribeiro, A.: Large scale wireless power allocation with graph neural networks. In: 2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC). pp. 1–5. IEEE (2019)
14. Mao, Y., Zhang, J., Letaief, K.B.: Dynamic computation offloading for mobile-edge computing with energy harvesting devices. IEEE Journal on Selected Areas in Communications 34(12), 3590–3605 (2016)
15. Xi, Y., Burr, A., Wei, J., Grace, D.: A general upper bound to evaluate packet error rate over quasi-static fading channels. IEEE Transactions on Wireless Communications 10(5), 1373–1377 (2011)
16. Zheng, J., Li, K., Tovar, E., Guizani, M.: Federated learning for energy-balanced client selection in mobile edge computing. In: 2021 International Wireless Communications and Mobile Computing (IWCMC). pp. 1942–1947. IEEE (2021)
17. Chen, T., Kim, H., Yang, Y.: Energy efficiency metrics for green wireless communications. In: 2010 International Conference on Wireless Communications & Signal Processing (WCSP). pp. 1–6. IEEE (2010)
18. Lee, J.B., Rossi, R.A., Kim, S., Ahmed, N.K., Koh, E.: Attention models in graphs: A survey. ACM Transactions on Knowledge Discovery from Data (TKDD) 13(6), 1–25 (2019)
19. Gama, F., Bruna, J., Ribeiro, A.: Stability properties of graph neural networks. IEEE Transactions on Signal Processing 68, 5680–5695 (2020)
20. Boyd, S., Vandenberghe, L.: Convex optimization. Cambridge university press (2004)
21. Yates, R.D., Sun, Y., Brown, D.R., Kaul, S.K., Modiano, E., Ulukus, S.: Age of information: An introduction and survey. IEEE Journal on Selected Areas in Communications 39(5), 1183–1210 (2021)
22. Wu, B., Fang, F., Wang, X.: Joint age-based client selection and resource allocation for communication-efficient federated learning over noma networks. IEEE Transactions on Communications (2023)
23. Eisen, M., Zhang, C., Chamon, L.F., Lee, D.D., Ribeiro, A.: Learning optimal resource allocations in wireless systems. IEEE Transactions on Signal Processing 67(10), 2775–2790 (2019)
24. Morrison, D.R., Jacobson, S.H., Sauppe, J.J., Sewell, E.C.: Branch-and-bound algorithms: A survey of recent advances in searching, branching, and pruning. Discrete Optimization 19, 79–102 (2016)
25. Calcev, G., Chizhik, D., Goransson, B., Howard, S., Huang, H., Kogiantis, A., Molisch, A.F., Moustakas, A.L., Reed, D., Xu, H.: A wideband spatial channel model for system-wide simulations. IEEE Transactions on Vehicular Technology 56(2), 389–403 (2007)

26. McMahan, B., Moore, E., Ramage, D., Hampson, S., y Arcas, B.A.: Communication-efficient learning of deep networks from decentralized data. In: Artificial intelligence and statistics. pp. 1273–1282. PMLR (2017)