SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

July 2025, Vol. 68, Iss. 7, 170304:1–170304:17
https://doi.org/10.1007/s11432-024-4433-9

Special Topic: Integration of Large AI Model and 6G

FAS-assisted federated learning over wireless communication systems†

Hao XU[1*], Kai-Kit WONG[2], Yongxu ZHU[1], Chongwen HUANG[3], Chao WANG[4], Wee Kiat NEW[2], Farshad Rostami GHADI[2] & Gui ZHOU[5]

[1]*National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China*
[2]*Department of Electronic and Electrical Engineering, University College London, London WC1E7JE, UK*
[3]*College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*
[4]*Integrated Service Networks Laboratory, Xidian University, Xi'an 710071, China*
[5]*Institute for Digital Communications, Friedrich-Alexander-University Erlangen-Nürnberg, Erlangen 91054, Germany*

Received 13 November 2024/Revised 20 February 2025/Accepted 8 May 2025/Published online 20 June 2025

Abstract This paper examines the energy efficiency of a multi-user system where federated learning (FL) is implemented in a distributed manner across all nodes. Each user employs a fluid antenna system (FAS) to improve the channel condition, while the base station (BS) is equipped with multiple traditional fixed-position antennas (FPAs). When performing the FL algorithm, each user first trains a local model and transmits it to the BS over shared time-frequency resources. Then, the BS aggregates the received models and broadcasts the combined model back to all users. These steps are repeated until the FL model achieves a desired accuracy level. The system energy is mainly consumed in the computation and transmission processes at the user side. To save energy, we develop an optimization framework that minimizes the total energy consumption by jointly optimizing the learning accuracy, transmit power, antenna positions, and the BS receivers. Since the optimization variables are highly coupled, the problem is non-convex and quite complex. To address the issue, we propose an iterative algorithm to obtain a suboptimal solution of the problem. Simulation results have verified the effectiveness of the algorithm and also the advantages of FAS over the conventional FPA technology.

Keywords fluid antenna system (FAS), federated learning (FL), multiple access, energy efficiency

Citation Xu H, Wong K-K, Zhu Y X, et al. FAS-assisted federated learning over wireless communication systems. Sci China Inf Sci, 2025, 68(7): 170304, https://doi.org/10.1007/s11432-024-4433-9

1 Introduction

The upcoming sixth-generation (6G) wireless networks are expected to provide extremely high capacity, reliability, massive connectivity, and services beyond communications. In line with this trend, fluid antenna system (FAS), representing any software-controllable fluidic, dielectric or conductive structures, such as mechanical liquid-based antennas, radio-frequency pixel-based antennas, movable antennas, or metasurface, that can reconfigure the shape, size, position, orientation, and other radiation characteristics, has been proposed in recent years for enabling flexible and adaptive wireless communications [1–5]. Unlike the traditional antenna techniques where multiple antennas are discretely deployed at fixed positions with sufficient separation, the very fine spatial resolution and dynamic shape of FAS enable it to capitalize on the full range of spatial variations, resulting in significantly improved performance in terms of outage probability [6–8], diversity gain [9, 10], system capacity [11–13], and secrecy rate [14, 15]. FAS also offers a new capability to exploit the spatial opportunity where the interference suffers from deep fades for multiuser communication, leading to the formation of new multiple access [16–18].

The effectiveness of FAS is closely tied to the accurate acquisition of channel state information (CSI). Therefore, the development of methods for obtaining accurate CSI at different positions or at a wide range of preset ports on a FAS is critical to realizing its full potential. Channel estimation for FAS-assisted systems has been studied by some studies [19, 20]. Specifically, in [19], a low-sample size method was proposed to estimate and reconstruct the CSI of a mmWave uplink system, where all transmitters

* Corresponding author (email: hao.xu@seu.edu.cn)
† Special Topic: The 27th Annual Meeting of the China Association for Science and Technology

use FAS. In [20], channel estimation and reconstruction were achieved through a combination of Nyquist sampling and maximum likelihood estimation. The study highlighted the importance of oversampling to improve estimation accuracy. However, given the practical limitations of oversampling, a suboptimal sampling distance was proposed to enable efficient channel reconstruction while balancing accuracy and feasibility. In addition to the above-mentioned papers, a number of studies have also explored how FAS can work synergistically with other emerging technologies, such as enhancing next-generation multiple access (NGMA) applications [21–23], advancing AI-enabled FAS systems [24–26], investigating FAS-assisted integrated sensing and communication (ISAC) [27–29], and facilitating communication with FAS and reconfigurable intelligent surface (RIS) integration [30, 31].

On the other hand, machine learning will be essential in next-generation mobile communication systems, providing the foundation for network intelligence [32–34]. However, due to privacy concerns, limited computational resources, and communication constraints, centralized learning, where the users transmit all of their collected data to a central base station (BS) for processing, is often impractical [35–38]. In response, distributed learning frameworks like federated learning (FL) have emerged as key solutions, allowing users to collaboratively train models while preserving data privacy through local data processing [39, 40]. The performance of FL in wireless systems has been extensively studied in several studies [41–43]. In particular, Ref. [41] addressed the FL loss function minimization problem, factoring in packet errors over wireless links. Studies in [42, 43] focused on minimizing the total energy consumed in computation and communication when executing the FL algorithms. However, Ref. [42] required that all users transmit their local model synchronously. To avoid co-channel interference, Refs. [41, 43] employed orthogonal frequencies for user transmissions, which may constrain the communication performance during model uploads. Moreover, all these studies considered the traditional fixed-position antennas (FPAs).

In this paper, we examine energy-efficient computation and transmission for FL in a FAS-assisted mobile system. Each user has a single-antenna FAS, while the BS is equipped with multiple traditional FPAs. Leveraging collected data, each user trains a local FL model using their own computational resources. After training, users transmit their models to the BS over shared time-frequency resources. The BS aggregates the received models and broadcasts the combined model back to all users. These steps, i.e., local training, data transmission, aggregation, and broadcasting, are repeated iteratively until the FL model achieves a desired accuracy level. The system's energy consumption depends mainly on the learning accuracy and data transmission size. To minimize overall energy usage, we develop an optimization framework that jointly optimizes the learning accuracy, transmit power, antenna positions, and the BS receivers. Due to the strong interdependence among variables, the problem is non-convex and quite intractable. To address this, we develop an iterative algorithm to deal with the problem, where in each iteration, the optimal or near-optimal solutions for the arrived subproblems can be easily found. Simulation results show that the proposed algorithm can greatly reduce the energy consumption, and also highlight the advantages of FAS over the traditional FPA technology.

The rest of this paper is organized as follows. In Section 2, the system model is provided and the problem is formulated. In Section 3, we provide an iterative algorithm to solve the energy minimization problem. Simulation results are given in Section 4 to verify the proposed algorithm. Finally, Section 5 draws some concluding remarks. Auxiliary technical results are given in the appendices.

Notations. $\mathbb{R}$ and $\mathbb{C}$, respectively, represent the real and complex spaces. Boldface lower and upper case letters are used to denote vectors and matrices. $\boldsymbol{I}_M$ stands for the $M \times M$ dimensional identity matrix and $\mathbf{0}$ denotes the all-zero vector or matrix. Also, superscripts $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$, respectively, denote the transpose and conjugate transpose operations. We use calligraphic capital letters to denote sets, such as $\mathcal{U} = \{1, \dots, U\}$, and $\mathcal{U}_1 \setminus \mathcal{U}_2$ to denote set subtraction. We use calligraphic subscript to denote the set of elements whose indexes take values from the subscript set, e.g., $\boldsymbol{p}_{\mathcal{U}} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_U\}$. If $|\cdot|$ is applied to a number, it denotes the magnitude of that number; when applied to a set, it represents the cardinality of the set. $\|\cdot\|_2$ is the Euclidean norm. To help readers follow the mathematical contents, the meanings of some key notations in this paper are summarized in Table 1.

2 System model and problem formulation

As depicted in Figure 1, we consider a multi-user wireless communication system with U users and a BS. Each user is equipped with a linear FAS of size W_u, where the antenna can be instantly switched to

Table 1 List of key notations.

Notation	Definition	Notation	Definition
$U, \mathcal{U}, \mathcal{U}_n$	Number of users, set of all users, set of active users at time block n	δ	Step size of the gradient method in computing the local FL model
$N, \mathcal{N}, \mathcal{N}_u$	Number of time blocks, set of all time blocks, set of active time blocks of user u	γ	γ-strongly convex
$W_u, w_{u,n}$	FAS size of user u, antenna position of the FAS of user u at time block n	κ	Effective switched capacitance in the local computation
M	Number of antennas at the BS	s	Transmit data size of the users
d	Spacing of the BS antenna ULA	$x_{u,n}, p_{u,n}$	Siganl and transmit power of user u at time blodk n
λ	Wavelength	$\boldsymbol{z}_n, \sigma^2$	Noise vector and noise power
B	Bandwidth	$\boldsymbol{y}_n$	Received signal vector of the BS at time block n
T	Time constraint on local computation and data transmission	$\boldsymbol{g}_{u,n}$	Channel vector from user u to the BS at time block n
τ_n	Length of time block n	L_u	Number of propagation paths between user u and the BS
C_u	Number of CPU cycles necessary for computing one sample data at user u	$\phi_{u,l}$	Complex channel gain of the l-th path of user u
D_u	Number of data samples of user u	$\beta_{u,l}, \theta_{u,l}$	AoA and AoD of the l-th path of user u
G_u	Computation capacity of user u	$\boldsymbol{a}_{u,\mathrm{R}}$	Steering vector at the receiver side
I_u	Number of computation iterations of user u	$a_{u,\mathrm{T}}$	Steering element at the transmitter side
J	Lipschitz constant for the loss function of the FL model	$\boldsymbol{f}_{u,n}$	Linear receiver for detecting $x_{u,n}$
η, ϵ_0	Local and global accuracy levels of the FL model	ρ	Fraction of time allocated for local FL model computation

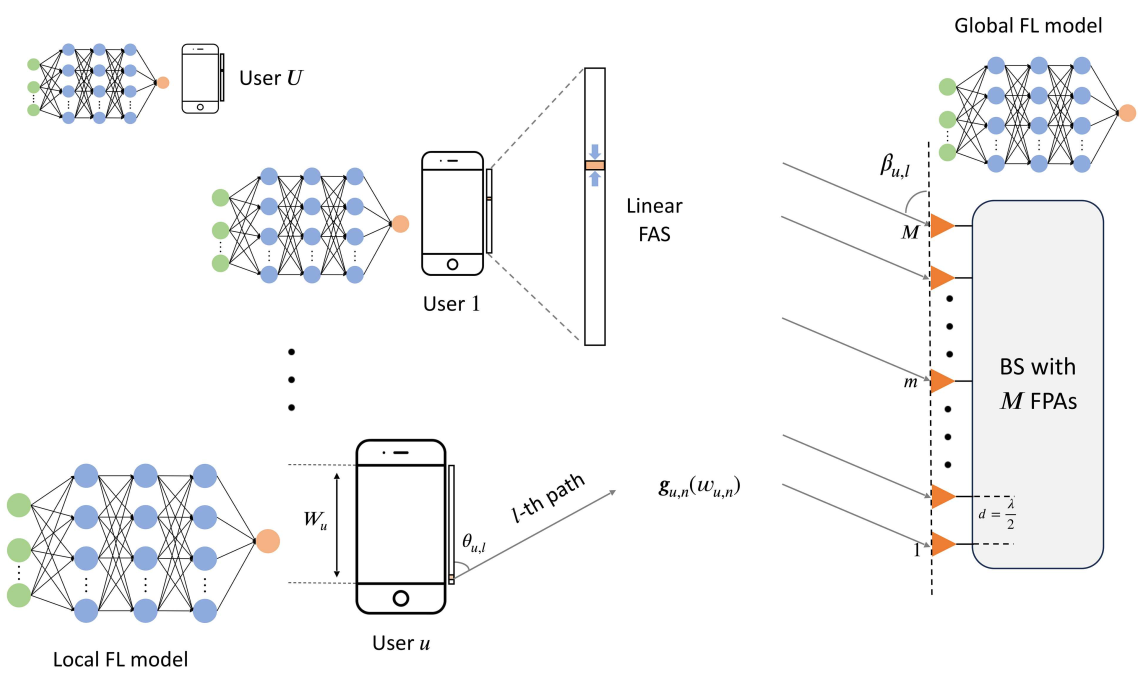

Figure 1 (Color online) FAS-assisted mobile system where the users have FAS for transmission but the BS uses multiple FPAs for reception.

any position in a FAS[1)] [44]. The BS is equipped with an M-antenna uniform linear array (ULA) with antenna spacing $d = \frac{\lambda}{2}$, where λ is the wavelength. All users and the BS collaboratively execute an FL algorithm across the network for data analysis and inference. The model trained at each user using its local dataset is referred to as the local FL model, while the model aggregated and updated at the BS using all collected local models is called the global FL model. The FL process is conducted iteratively between the users and the BS, with each iteration consisting of three main steps: local computation at

1) Each antenna structure can be seen as a reconfigurable pixel-based linear FAS that has many tiny pixels, among which the certain number of pixels and their connections are activated at each time to switch between many reconfigurable states [44]. Using this technology, it is possible to switch the antennas with almost no time delay.

each user, local FL model transmission to the BS, and aggregation of results followed by broadcast from the BS.

Note that in the downlink phase, the BS benefits from high power and bandwidth for broadcasting the global FL model. Thus, as assumed in [43], the downlink transmission time is considered negligible compared to the uplink transmission time. In the following, we first introduce the computation and transmission models, and then formulate the energy-efficiency optimization problem.

2.1 Computation model

In the system, each user u has to compute a local dataset with D_u data samples. Denote the computation capacity of user u by G_u (cycles/second), which is measured by the number of CPU cycles per second. Then, the time of user u required for computing its local FL model is

$$t_u^{\mathrm{C}} = \frac{I_u C_u D_u}{G_u}, \tag{1}$$

where C_u (cycles/sample) is the number of CPU cycles necessary for computing one sample data at user u. In addition, I_u in (1) is the number of computation iterations of user u and is given by

$$I_u = v \log_2 \frac{1}{\eta}, \tag{2}$$

where η is the local accuracy level and $v = \frac{2}{(2-J\delta)\delta\gamma}$ [43, Lemma 1]. Here J is the Lipschitz constant for the loss function of the FL model, $\gamma > 0$ is a real positive number that indicates that the loss function of the FL model is γ-strongly convex, and $\delta < \frac{2}{J}$ is the step size of the gradient method used by each user in computing the local FL model[2)] [43]. Then, according to [45, Lemma 1], the energy required to compute a total of $C_u D_u$ CPU cycles over I_u iterations is

$$E_u^{\mathrm{C}} = \kappa I_u C_u D_u G_u^2, \tag{3}$$

where κ is the effective switched capacitance that depends on the chip architecture.

2.2 Transmission model

In the uplink transmission phase, the users communicate with the BS using the same time-frequency resource with bandwidth B. Assume that the users must complete their local computation and data transmission within a total time of T s. As illustrated in Figure 2, user u spends t_u^{C} on local computation, leaving $t_u^{\mathrm{T}} = T - t_u^{\mathrm{C}}$ for data transmission. Since local computations may take different amounts of time for the users, transmissions may thus start at different times for them. Obviously, the user completing its local computation in the shortest time will begin transmission first, while the user with the longest local computation time will begin transmission last. We define the interval $T - \min\{t_u^{\mathrm{C}}, \ldots, t_U^{\mathrm{C}}\}$ as the transmission time for the user that completes the computation earliest. This interval is divided into N time blocks $\tau_1, \ldots, \tau_N$, with each block marking the start of transmission for a new user who has just finished local computation and begins uploading its FL model. Since there are in total U users, we have at most U transmission time blocks, i.e., $N \leqslant U$. For convenience, let $\mathcal{U}_n$ represent the set of users transmitting their FL models during time block n, and let $\mathcal{N}_u$ denote the set of time block indexes in which user u transmits it FL model.

For ease of understanding, we take the system with three users in Figure 2 for example. Since $t_2^{\mathrm{C}} < t_1^{\mathrm{C}} < t_3^{\mathrm{C}}$, the interval $t_2^{\mathrm{T}} = T - t_2^{\mathrm{C}}$ is divided into three blocks, i.e., $\tau_1 = t_1^{\mathrm{C}} - t_2^{\mathrm{C}}$, $\tau_2 = t_3^{\mathrm{C}} - t_1^{\mathrm{C}}$, and $\tau_1 = T - t_3^{\mathrm{C}}$. During the first time block τ_1, only user 2 is transmitting. Therefore, $\mathcal{U}_1 = \{2\}$. Similarly, we have $\mathcal{U}_2 = \{1, 2\}$ and $\mathcal{U}_3 = \{1, 2, 3\}$. On the other hand, user 1 transmits its FL model in the second and third time blocks. Therefore, $\mathcal{N}_1 = \{2, 3\}$. Analogously, we have $\mathcal{N}_2 = \{1, 2, 3\}$ and $\mathcal{N}_3 = \{3\}$.

2) Note that in this paper, we adopt the same FL algorithm as used in [43]. Therefore, several parameters from [43], such as J, γ, and δ, are also relevant here. Due to space limitations, we do not detail the computations for the local and global FL models. For further information on the FL algorithm and the specific meanings of these parameters, please refer to [43].

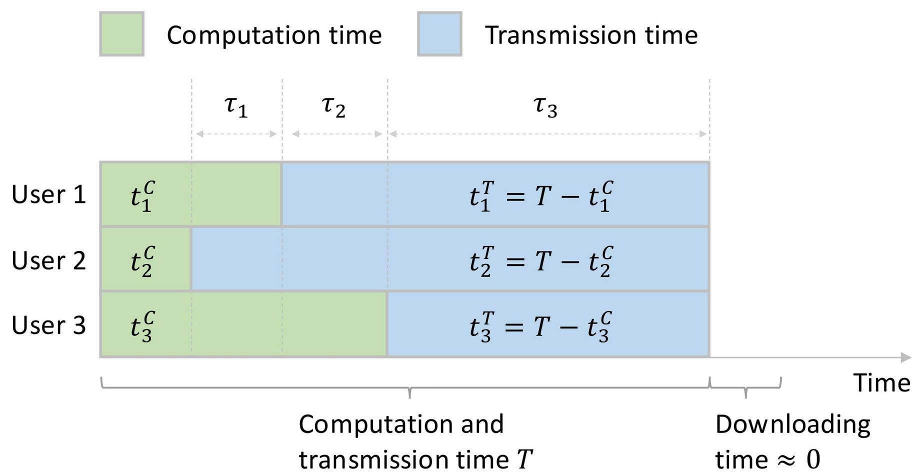

Figure 2 (Color online) Implementation of the FL algorithm with three users.

2.2.1 *Signal model*

As stated above, at the n-th time block, users in the set $\mathcal{U}_n$ are transmitting their FL models simultaneously using the shared time-frequency resource. Then, the received signal of the BS at this time block is given by

$$\boldsymbol{y}_n = \sum_{u \in \mathcal{U}_n} \boldsymbol{g}_{u,n} x_{u,n} + \boldsymbol{z}_n, \tag{4}$$

where $x_{u,n} \sim \mathcal{CN}(0, p_{u,n})$ is the signal of user u, $p_{u,n}$ is the transmit power, $\boldsymbol{g}_{u,n} \in \mathbb{C}^{M \times 1}$ denotes the channel vector from user u to the BS, and $\boldsymbol{z}_n \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$ represents the additive white Gaussian noise (AWGN).

2.2.2 *Channel model*

Let $w_{u,n} \in [0, W_u]$ denote the antenna position of user u at the n-th time block. Using the planar-wave geometric channel model typical in mmWave systems [46], the channel vector $\boldsymbol{g}_{u,n}$ can be modeled as

$$\boldsymbol{g}_{u,n} = \sqrt{M} \sum_{l=1}^{L_u} \phi_{u,l} \boldsymbol{a}_{u,\mathrm{R}}(\beta_{u,l}) a_{u,\mathrm{T}}^*(\theta_{u,l}, w_{u,n}), \tag{5}$$

where L_u is the number of propagation paths between user u and the BS, and $\phi_{u,l} \sim \mathcal{CN}(0, 10^{-\frac{\mathrm{PL}_{u,l}}{10}})$ is the complex channel gain of the l-th path. Here $\mathrm{PL}_{u,l}$ is the distance-dependent pathloss of the l-th path of user u and it is given in the simulation part. Also, $\boldsymbol{a}_{u,\mathrm{R}}(\beta_{u,l})$ and $a_{u,\mathrm{T}}(\theta_{u,l}, w_{u,n})$ are, respectively, the steering vector at the receiver side and the steering element at the transmitter side, given by

$$\boldsymbol{a}_{u,\mathrm{R}}(\beta_{u,l}) = \frac{1}{\sqrt{M}} \left[1, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} d \cos \beta_{u,l}}, \ldots, \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda}(M-1) d \cos \beta_{u,l}}\right]^{\mathrm{T}}, \tag{6}$$

$$a_{u,\mathrm{T}}(\theta_{u,l}, w_{u,n}) = \mathrm{e}^{-\mathrm{j}\frac{2\pi}{\lambda} w_{u,n} \cos \theta_{u,l}}, \tag{7}$$

where $\beta_{u,l}, \theta_{u,l} \in [0, \pi]$ are, respectively, the angle-of-arrival (AoA) and angle-of-departure (AoD) of the l-th path of user u. Now, we define

$$\begin{aligned} \boldsymbol{\Phi}_u &= \mathrm{diag}\{\phi_{u,1}, \ldots, \phi_{u,L_u}\} \in \mathbb{C}^{L_u \times L_u}, \\ \boldsymbol{A}_{u,\mathrm{R}} &= [\boldsymbol{a}_{u,\mathrm{R}}(\beta_{u,1}), \ldots, \boldsymbol{a}_{u,\mathrm{R}}(\beta_{u,L_u})] \in \mathbb{C}^{M \times L_u}, \\ \boldsymbol{a}_{u,\mathrm{T}}(w_{u,n}) &= [a_{u,\mathrm{T}}(\theta_{u,1}, w_{u,n}), \ldots, a_{u,\mathrm{T}}(\theta_{u,L_u}, w_{u,n})] \in \mathbb{C}^{1 \times L_u}, \end{aligned} \tag{8}$$

based on which the channel vector $\boldsymbol{g}_{u,n}$ in (5) can also be expressed as

$$\boldsymbol{g}_{u,n} = \sqrt{M} \boldsymbol{A}_{u,\mathrm{R}} \boldsymbol{\Phi}_u \boldsymbol{a}_{u,\mathrm{T}}^{\mathrm{H}}(w_{u,n}). \tag{9}$$

2.2.3 *Wireless transmission*

Based on the above settings, the total energy consumed by user u for transmitting its FL model is

$$E_u^{\mathrm{T}} = \sum_{n\in\mathcal{N}_u} \tau_n p_{u,n}. \tag{10}$$

Assume that the BS uses a linear receiver $\boldsymbol{f}_{u,n} \in \mathbb{C}^{M\times 1}$ to detect the symbol $x_{u,n}$. Then, according to (4), the data rate of user u at the n-th time block is

$$R_{u,n} = \log_2\left(1 + \mathrm{SINR}_{u,n}\right), \tag{11}$$

where the signal-to-interference-plus-noise ratio (SINR) is given by

$$\mathrm{SINR}_{u,n} = \frac{p_{u,n}\left|\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u,n}\right|^2}{\sum_{u'\in\mathcal{U}_n\setminus\{u\}} p_{u',n}\left|\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u',n}\right|^2 + \sigma^2\left\|\boldsymbol{f}_{u,n}\right\|_2^2}. \tag{12}$$

Assuming the dimensions of the users' local FL models are fixed and equal, the data size that each user must upload is thus a constant, denoted by s. Each user u uploads its data over the time blocks in $\mathcal{N}_u$. Therefore, the following constraint has to be satisfied:

$$\sum_{n\in\mathcal{N}_u} B\tau_n R_{u,n} \geqslant s,\ \forall u\in\mathcal{U}. \tag{13}$$

2.3 Problem formulation

From (3) and (10) we know that the total energy consumption required for local FL model computation and transmission is

$$E = I_0 \sum_{u\in\mathcal{U}} \left(E_u^{\mathrm{C}} + E_u^{\mathrm{T}}\right), \tag{14}$$

where $I_0 = \frac{b}{1-\eta}$ is the number of global iterations in performing the FL algorithm and $b = \frac{2J^2}{\gamma^2\xi}\ln\frac{1}{\epsilon_0}$ [43, Theorem 1]. Here ξ is a constant value and ϵ_0 denotes the target accuracy level that the global FL model must achieve. Our objective is to minimize E by optimizing the local accuracy level, users' transmit power, antenna positions, and the BS's linear receivers. The problem can be formulated as

$$\begin{aligned}
\min_{\eta,\boldsymbol{p}_{\mathcal{U}},\boldsymbol{w}_{\mathcal{U}},\boldsymbol{F}_{\mathcal{U}}}\quad & E && \text{(15a)}\\
\text{s.t.}\quad & t_u^{\mathrm{C}} \leqslant \rho T,\ \forall u\in\mathcal{U}, && \text{(15b)}\\
& 0 \leqslant \eta \leqslant 1, && \text{(15c)}\\
& \sum_{n\in\mathcal{N}_u} B\tau_n R_{u,n} \geqslant s,\ \forall u\in\mathcal{U}, && \text{(15d)}\\
& p_{u,n} \geqslant 0,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, && \text{(15e)}\\
& 0 \leqslant w_{u,n} \leqslant W_u,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, && \text{(15f)}\\
& p_{u,n} = w_{u,n} = 0,\ \boldsymbol{f}_{u,n} = \boldsymbol{0},\ \forall u\in\mathcal{U},\ n\in\mathcal{N}\setminus\mathcal{N}_u, && \text{(15g)}
\end{aligned}$$

where $0 < \rho < 1$ represents the fraction of time allocated for local FL model computation, $\boldsymbol{p}_{\mathcal{U}} = \{\boldsymbol{p}_1,\ldots,\boldsymbol{p}_U\}$, $\boldsymbol{p}_u = [p_{u,1},\ldots,p_{u,N}]^{\mathrm{T}}$, $\boldsymbol{w}_{\mathcal{U}} = \{\boldsymbol{w}_1,\ldots,\boldsymbol{w}_U\}$, $\boldsymbol{w}_u = [w_{u,1},\ldots,w_{u,N}]^{\mathrm{T}}$, $\boldsymbol{F}_{\mathcal{U}} = \{\boldsymbol{F}_1,\ldots,\boldsymbol{F}_U\}$ and $\boldsymbol{F}_u = [\boldsymbol{f}_{u,1},\ldots,\boldsymbol{f}_{u,N}]$. Note that in (15g), the transmit power, antenna position, and BS receiver of user u are set to zero for any time block $n\in\mathcal{N}\setminus\mathcal{N}_u$. This is because user u only transmits data during the time blocks within $\mathcal{N}_u$.

3 Iterative algorithm

Since the optimization variables are highly interdependent, problem (15) is quite intractable. In this section, we propose an iterative algorithm to solve it.

3.1 Optimization of η

First, we optimize η with all the other variables fixed. Problem (15) in this case reduces to

$$\min_{\eta} \quad E \tag{16a}$$

$$\text{s.t.} \quad t_u^{\mathrm{C}} \leqslant \rho T, \ \forall u \in \mathcal{U}, \tag{16b}$$

$$0 \leqslant \eta \leqslant 1. \tag{16c}$$

To make the analysis more convenient, we first simplify the expression of (16). Let

$$\begin{aligned} \alpha_1 &= \frac{b\kappa v}{\ln 2} \sum_{u\in\mathcal{U}} C_u D_u G_u^2, \\ \alpha_2 &= b \sum_{u\in\mathcal{U}} \sum_{n\in\mathcal{N}_u} \tau_n p_{u,n}. \end{aligned} \tag{17}$$

Then, based on (2)–(10), the total energy consumption E in (14) can be rewritten as

$$\begin{aligned} E &= \frac{b}{1-\eta} \sum_{u\in\mathcal{U}} \left(\kappa v C_u D_u G_u^2 \log_2 \frac{1}{\eta} + \sum_{n\in\mathcal{N}_u} \tau_n p_{u,n} \right) \\ &= \frac{-\alpha_1 \ln \eta + \alpha_2}{1-\eta}. \end{aligned} \tag{18}$$

In addition, according to (1) and (2), the constraint (16b) can be rewritten as

$$-\frac{vC_u D_u}{G_u \ln 2} \ln \eta \leqslant \rho T, \ \forall u \in \mathcal{U}, \tag{19}$$

which can be equivalently transformed to

$$\eta \geqslant \alpha_0, \tag{20}$$

where

$$\alpha_0 = \max \left\{ \exp\left(-\frac{\rho T G_u \ln 2}{v C_u D_u} \right), \ \forall u \in \mathcal{U} \right\}. \tag{21}$$

Since $\frac{\rho T G_u \ln 2}{v C_u D_u} \geqslant 0, \forall u \in \mathcal{U}$, we know that $0 \leqslant \alpha_0 \leqslant 1$. Then, problem (16) can be equivalently transformed to

$$\min_{\eta} \quad \frac{-\alpha_1 \ln \eta + \alpha_2}{1-\eta} \tag{22a}$$

$$\text{s.t.} \quad \alpha_0 \leqslant \eta \leqslant 1. \tag{22b}$$

In Theorem 1, we provide the optimal solution of (22).

Theorem 1. Let η_0 denote the zero point of $\frac{1}{\eta} + \ln \eta = 1 + \frac{\alpha_2}{\alpha_1}$, which is unique and can be easily found using the bisection method. Then, the optimal solution of (22) is

$$\eta^* = \max\{\alpha_0, \eta_0\}. \tag{23}$$

Proof. See Appendix A.

3.2 Optimization of $\boldsymbol{p}_{\mathcal{U}}$, $\boldsymbol{w}_{\mathcal{U}}$, and $\boldsymbol{F}_{\mathcal{U}}$

Next, we optimize $\boldsymbol{p}_{\mathcal{U}}$, $\boldsymbol{w}_{\mathcal{U}}$, and $\boldsymbol{F}_{\mathcal{U}}$ for given η. The original problem (15) in this case reduces to

$$\min_{\boldsymbol{p}_{\mathcal{U}}, \boldsymbol{w}_{\mathcal{U}}, \boldsymbol{F}_{\mathcal{U}}} \quad \sum_{u\in\mathcal{U}} \sum_{n\in\mathcal{N}_u} \tau_n p_{u,n} \tag{24a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{N}_u} B \tau_n R_{u,n} \geqslant s, \ \forall u \in \mathcal{U}, \tag{24b}$$

$$p_{u,n} \geqslant 0,\ \forall u \in \mathcal{U},\ n \in \mathcal{N}_u, \tag{24c}$$

$$0 \leqslant w_{u,n} \leqslant W_u,\ \forall u \in \mathcal{U},\ n \in \mathcal{N}_u, \tag{24d}$$

$$p_{u,n} = w_{u,n} = 0,\ \boldsymbol{f}_{u,n} = \boldsymbol{0},\ \forall u \in \mathcal{U},\ n \in \mathcal{N} \setminus \mathcal{N}_u, \tag{24e}$$

where in the objective function, we omit the constants E_u^{C} and I_0 for convenience. From the channel model in (9) and the SINR expression in (12), we know that the optimization variables are coupled, making problem (24) non-convex and quite intractable. To address this issue, we use the fact that if minimum mean square error (MMSE) receivers are adopted by the BS for signal detection, the MMSE and SINR of each link satisfy $\mathrm{MMSE} = \frac{1}{1+\mathrm{SINR}}$ [47, 48]. This relationship allows us to simplify problem (24). In particular, the mean square error (MSE) of user u at time block n is given by

$$\begin{aligned}
\mathrm{MSE}_{u,n} &= \mathbb{E}\left[\left|\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{y} - \frac{1}{\sqrt{p_{u,n}}}x_{u,n}\right|^2\right] \\
&= \mathbb{E}\left[\left|\left(\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u,n} - \frac{1}{\sqrt{p_{u,n}}}\right)x_{u,n} + \sum_{u' \in \mathcal{U}_n \setminus \{u\}} \boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u',n}x_{u',n} + \boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{z}\right|^2\right] \\
&= \left(\sqrt{p_{u,n}}\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u,n} - 1\right)^2 + \sum_{u' \in \mathcal{U}_n \setminus \{u\}} p_{u',n}\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}}\boldsymbol{f}_{u,n} + \sigma^2\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{f}_{u,n} \\
&= \boldsymbol{f}_{u,n}^{\mathrm{H}}\left(\sum_{u' \in \mathcal{U}_n} p_{u',n}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_M\right)\boldsymbol{f}_{u,n} - \sqrt{p_{u,n}}\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u,n} - \sqrt{p_{u,n}}\boldsymbol{g}_{u,n}^{\mathrm{H}}\boldsymbol{f}_{u,n} + 1,
\end{aligned} \tag{25}$$

and the corresponding MMSE receiver that minimizes $\mathrm{MSE}_{u,n}$ is

$$\begin{aligned}
\boldsymbol{f}_{u,n}^* &= \min_{\boldsymbol{f}_{u,n}} \mathrm{MSE}_{u,n} \\
&= \sqrt{p_{u,n}}\left(\sum_{u' \in \mathcal{U}_n} p_{u',n}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_M\right)^{-1}\boldsymbol{g}_{u,n}.
\end{aligned} \tag{26}$$

Since $\mathrm{MMSE}_{u,n} = \frac{1}{1+\mathrm{SINR}_{u,n}}$, problem (24) can be transformed to

$$\min_{\boldsymbol{p}_{\mathcal{U}}, \boldsymbol{w}_{\mathcal{U}}, \boldsymbol{F}_{\mathcal{U}}} \quad \sum_{u \in \mathcal{U}} \sum_{n \in \mathcal{N}_u} \tau_n p_{u,n} \tag{27a}$$

$$\text{s.t.} \quad \sum_{n \in \mathcal{N}_u} \tau_n \ln \mathrm{MSE}_{u,n} \leqslant -\hat{s},\ \forall u \in \mathcal{U}, \tag{27b}$$

$$p_{u,n} \geqslant 0,\ \forall u \in \mathcal{U},\ n \in \mathcal{N}_u, \tag{27c}$$

$$0 \leqslant w_{u,n} \leqslant W_u,\ \forall u \in \mathcal{U},\ n \in \mathcal{N}_u, \tag{27d}$$

$$p_{u,n} = w_{u,n} = 0,\ \boldsymbol{f}_{u,n} = \boldsymbol{0},\ \forall u \in \mathcal{U},\ n \in \mathcal{N} \setminus \mathcal{N}_u, \tag{27e}$$

where $\hat{s} = \frac{s \ln 2}{B}$. Note that although the relationship $\mathrm{MMSE}_{u,n} = \frac{1}{1+\mathrm{SINR}_{u,n}}$ is adopted, we use $\mathrm{MSE}_{u,n}$ in (27b). Now we explain the reason. From the expressions of the channel vector $\boldsymbol{g}_{u,n}$ in (9) and the MMSE receiver $\boldsymbol{f}_{u,n}^*$ in (26), we know that the MMSE receiver minimizing $\mathrm{MSE}_{u,n}$ is a function of $p_{u,n}$ and $w_{u,n}$ for all $u \in \mathcal{U}_n$. While substituting $\boldsymbol{f}_{u,n}^*$ into (25) would allow us to reformulate problem (27) solely in terms of $\boldsymbol{p}_{\mathcal{U}}$ and $\boldsymbol{w}_{\mathcal{U}}$, this approach would result in strong coupling among the variables $p_{u,n}, w_{u,n}, \forall u \in \mathcal{U}_n$, rendering problem (27) significantly more intractable. Therefore, we retain $\boldsymbol{f}_{u,n}$ as an independent variable such that $\boldsymbol{p}_{\mathcal{U}}$, $\boldsymbol{w}_{\mathcal{U}}$, and $\boldsymbol{F}_{\mathcal{U}}$ can be optimized in an alternative way. Since there is no constraint on the value of $\boldsymbol{f}_{u,n}$, we use MSE in (27b). However, it can be easily proven that in the optimal case of (27), $\boldsymbol{f}_{u,n}$ must adopt the form given in (26), thereby transforming the MSE into the MMSE.

Although the fractional SINR expressions have been avoided, problem (27) remains challenging since the left-hand-side term of each constraint (27b) is a weighted sum of $|\mathcal{N}_u|$ $\ln(\cdot)$ terms. To remove the $\ln(\cdot)$ operation, we introduce an auxiliary variable $\mu_{u,n}$ for each user u and time block n, and arrive at a new problem as follows:

$$\min_{\boldsymbol{p}_{\mathcal{U}},\boldsymbol{w}_{\mathcal{U}},\boldsymbol{F}_{\mathcal{U}},\boldsymbol{\mu}_{\mathcal{U}}} \quad \sum_{u\in\mathcal{U}}\sum_{n\in\mathcal{N}_u}\tau_n p_{u,n} \tag{28a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{N}_u}\tau_n\left(\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}_{u,n}-\mu_{u,n}\right)\leqslant -\hat{s},\ \forall u\in\mathcal{U}, \tag{28b}$$

$$p_{u,n}\geqslant 0,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, \tag{28c}$$

$$0\leqslant w_{u,n}\leqslant W_u,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, \tag{28d}$$

$$p_{u,n}=w_{u,n}=\mu_{u,n}=0,\ \boldsymbol{f}_{u,n}=\boldsymbol{0},\ \forall u\in\mathcal{U},\ n\in\mathcal{N}\setminus\mathcal{N}_u, \tag{28e}$$

where $\boldsymbol{\mu}_{\mathcal{U}}=\{\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_U\}$ and $\boldsymbol{\mu}_u=[\mu_{u,1},\ldots,\mu_{u,N}]^{\mathrm{T}}$. In Theorem 2, we show that the above two problems (27) and (28) are equivalent and provide the optimal solution for each $\mu_{u,n}$.

Theorem 2. The problems (27) and (28) are equivalent in that they have the same optimal objective function value. Let $(\boldsymbol{p}_{\mathcal{U}}^*,\boldsymbol{w}_{\mathcal{U}}^*,\boldsymbol{F}_{\mathcal{U}}^*,\boldsymbol{\mu}_{\mathcal{U}}^*)$ denote the optimal solution of (28) and $\mathrm{MSE}_{u,n}^*$ represent the corresponding MSE of user u at time block n. Then, we have

$$\mu_{u,n}^*=1-\ln\mathrm{MSE}_{u,n}^*,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u. \tag{29}$$

Proof. See Appendix B.

Note that although problem (28) has been significantly simplified, it remains complex. In the following, we solve it by iteratively optimizing the variables.

3.2.1 *Optimization of* $\boldsymbol{p}_{\mathcal{U}}$

We first optimize $\boldsymbol{p}_{\mathcal{U}}$ for given $\boldsymbol{w}_{\mathcal{U}}$, $\boldsymbol{F}_{\mathcal{U}}$, and $\boldsymbol{\mu}_{\mathcal{U}}$. For convenience, define $q_{u,n}=\sqrt{p_{u,n}},\forall u\in\mathcal{U},n\in\mathcal{N}$, $\boldsymbol{q}_u=[q_{u,1},\ldots,q_{u,N}]^{\mathrm{T}}$, and $\boldsymbol{q}_{\mathcal{U}}=\{\boldsymbol{q}_1,\ldots,\boldsymbol{q}_U\}$. Then, based on (25), problem (28) becomes

$$\min_{\boldsymbol{q}_{\mathcal{U}}} \quad \sum_{u\in\mathcal{U}}\sum_{n\in\mathcal{N}_u}\tau_n q_{u,n}^2 \tag{30a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{N}_u}\tau_n\left\{\mathrm{e}^{\mu_{u,n}-1}\left[\left(q_{u,n}\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u,n}-1\right)^2+\sum_{u'\in\mathcal{U}_n\setminus\{u\}}q_{u',n}^2\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}}\boldsymbol{f}_{u,n}+\sigma^2\boldsymbol{f}_{u,n}^{\mathrm{H}}\boldsymbol{f}_{u,n}\right]-\mu_{u,n}\right\}$$
$$\leqslant -\hat{s},\ \forall u\in\mathcal{U}, \tag{30b}$$

$$q_{u,n}\geqslant 0,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, \tag{30c}$$

$$q_{u,n}=0,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}\setminus\mathcal{N}_u. \tag{30d}$$

Problem (30) is a quadratically constrained quadratic program (QCQP) and is convex. The optimal solution can thus be found by using the standard tools such as Interior-point method and CVX. Once the optimal $\boldsymbol{q}_{\mathcal{U}}^*$ is obtained, we set $p_{u,n}^*=q_{u,n}^{*2},\forall u\in\mathcal{U},n\in\mathcal{N}$.

3.2.2 *Optimization of* $\boldsymbol{w}_{\mathcal{U}}$ *and* $\boldsymbol{F}_{\mathcal{U}}$

Next, we optimize $\boldsymbol{w}_{\mathcal{U}}$ and $\boldsymbol{F}_{\mathcal{U}}$ with given $\boldsymbol{p}_{\mathcal{U}}$ and $\boldsymbol{\mu}_{\mathcal{U}}$. From (28) we see that the values of $\boldsymbol{w}_{\mathcal{U}}$ and $\boldsymbol{F}_{\mathcal{U}}$ do not affect the objective function (28a), but only affect the MSE terms in the constraint (28b). Considering that minimizing the sum MSE of all users at all time blocks can provide more degree of freedom in optimizing the other variables in the subsequent steps, we optimize $\boldsymbol{w}_{\mathcal{U}}$ and $\boldsymbol{F}_{\mathcal{U}}$ by minimizing the summation of the left-hand-side terms of (28b) for all users. The problem can be formulated as follows:

$$\min_{\boldsymbol{w}_{\mathcal{U}},\boldsymbol{F}_{\mathcal{U}}} \quad \sum_{u\in\mathcal{U}}\sum_{n\in\mathcal{N}_u}\tau_n\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}_{u,n} \tag{31a}$$

$$\text{s.t.} \quad \sum_{n\in\mathcal{N}_u}\tau_n\left(\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}_{u,n}-\mu_{u,n}\right)\leqslant -\hat{s},\ \forall u\in\mathcal{U}, \tag{31b}$$

$$0\leqslant w_{u,n}\leqslant W_u,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u, \tag{31c}$$

$$w_{u,n}=0,\ \forall u\in\mathcal{U},\ n\in\mathcal{N}\setminus\mathcal{N}_u. \tag{31d}$$

From the channel model in Subsection 2.2.2 and (25) we observe that the variables $w_{u,n}$, $\boldsymbol{f}_{u,n}, \forall u \in \mathcal{U}, n \in \mathcal{N}_u$ are highly interdependent in the MSE expressions, making (31) quite challenging to solve. To make the problem tractable, we iteratively optimize the antenna position of each user at each time block, i.e., $w_{u,n}$. Note that although, as seen from (9), the variable $w_{u,n}$ directly influences only the channel vector $\boldsymbol{g}_{u,n}$, the resulting $\boldsymbol{g}_{u,n}$ subsequently affects the MMSE receivers of all users in the n-th time block, i.e., $\boldsymbol{f}_{u'',n}, \forall u'' \in \mathcal{U}_n$. Consequently, in each iteration, in addition to updating $w_{u,n}$, it is necessary to also update $\boldsymbol{f}_{u'',n}, \forall u'' \in \mathcal{U}_n$. Nevertheless, since $\boldsymbol{f}_{u'',n}, \forall u'' \in \mathcal{U}_n$ can be computed using the closed-form expression provided in (26), the updating process remains computationally efficient. The problem is formulated as follows:

$$\min_{w_{u,n},\{\boldsymbol{f}_{u,n}\}_{u''\in\mathcal{U}_n}} \quad \sum_{u''\in\mathcal{U}_n} \mathrm{e}^{\mu_{u'',n}-1}\mathrm{MSE}_{u'',n} \tag{32a}$$

$$\text{s.t.} \quad \sum_{n'\in\mathcal{N}_{u''}} \tau_{n'}\left(\mathrm{e}^{\mu_{u'',n'}-1}\mathrm{MSE}_{u'',n'} - \mu_{u'',n'}\right) \leqslant -\hat{s},\ \forall u'' \in \mathcal{U}_n, \tag{32b}$$

$$0 \leqslant w_{u,n} \leqslant W_u, \tag{32c}$$

where the parameter τ_n in the objective function is omitted for brevity since it is the weighted sum of the MSE of all users at the fixed time block n. Note that to differentiate from the specific user u and time block n under consideration, in problem (32), we introduce the notation u'' to represent a general user in $\mathcal{U}_n$ and n' for a general time block in $\mathcal{N}_{u''}$. In the following, we show that problem (32) can be near-optimally solved.

Before solving (32), we first simplify its expression. According to (25), for any user $u'' \in \mathcal{U}_n$, its MSE at the n-th time block can be written as

$$\begin{aligned}\mathrm{MSE}_{u'',n} &= \boldsymbol{f}_{u'',n}^{\mathrm{H}}\left(\sum_{u'\in\mathcal{U}_n} p_{u',n}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_M\right)\boldsymbol{f}_{u'',n} - \sqrt{p_{u'',n}}\boldsymbol{f}_{u'',n}^{\mathrm{H}}\boldsymbol{g}_{u'',n} - \sqrt{p_{u'',n}}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\boldsymbol{f}_{u'',n} + 1\\ &= \boldsymbol{f}_{u'',n}^{\mathrm{H}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)\boldsymbol{f}_{u'',n} - \sqrt{p_{u'',n}}\boldsymbol{f}_{u'',n}^{\mathrm{H}}\boldsymbol{g}_{u'',n} - \sqrt{p_{u'',n}}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\boldsymbol{f}_{u'',n} + 1\\ &= -p_{u'',n}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)^{-1}\boldsymbol{g}_{u'',n} + 1,\end{aligned} \tag{33}$$

where

$$\boldsymbol{Q}_{u,n} = \sum_{u'\in\mathcal{U}_n\setminus\{u\}} p_{u',n}\boldsymbol{g}_{u',n}\boldsymbol{g}_{u',n}^{\mathrm{H}} + \sigma^2\boldsymbol{I}_M, \tag{34}$$

and the last step of (33) holds since in the optimal case, $\boldsymbol{f}_{u'',n} = \sqrt{p_{u'',n}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)^{-1}\boldsymbol{g}_{u'',n}$. Using (33), the objective function (32a) can be rewritten as

$$\sum_{u''\in\mathcal{U}_n} \mathrm{e}^{\mu_{u'',n}-1}\mathrm{MSE}_{u'',n} = \sum_{u''\in\mathcal{U}_n} \mathrm{e}^{\mu_{u'',n}-1}\left[-p_{u'',n}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)^{-1}\boldsymbol{g}_{u'',n} + 1\right]. \tag{35}$$

In addition, the constraint (32b) can be rewritten as

$$\begin{aligned}&\sum_{n'\in\mathcal{N}_{u''}} \tau_{n'}\left(\mathrm{e}^{\mu_{u'',n'}-1}\mathrm{MSE}_{u'',n'} - \mu_{u'',n'}\right)\\ &= \tau_n\left(\mathrm{e}^{\mu_{u'',n}-1}\mathrm{MSE}_{u'',n} - \mu_{u'',n}\right) + \sum_{n'\in\mathcal{N}_{u''}\setminus\{n\}} \tau_{n'}\left(\mathrm{e}^{\mu_{u'',n'}-1}\mathrm{MSE}_{u'',n'} - \mu_{u'',n'}\right)\\ &= \tau_n\left\{\mathrm{e}^{\mu_{u'',n}-1}\left[-p_{u'',n}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)^{-1}\boldsymbol{g}_{u'',n} + 1\right] - \mu_{u'',n}\right\}\\ &\quad + \sum_{n'\in\mathcal{N}_{u''}\setminus\{n\}} \tau_{n'}\left(\mathrm{e}^{\mu_{u'',n'}-1}\mathrm{MSE}_{u'',n'} - \mu_{u'',n'}\right)\\ &\leqslant -\hat{s},\ \forall u'' \in \mathcal{U}_n.\end{aligned} \tag{36}$$

For convenience, we equivalently rewrite the inequality (36) as follows:

$$-p_{u'',n}\boldsymbol{g}_{u'',n}^{\mathrm{H}}\left(p_{u,n}\boldsymbol{g}_{u,n}\boldsymbol{g}_{u,n}^{\mathrm{H}} + \boldsymbol{Q}_{u,n}\right)^{-1}\boldsymbol{g}_{u'',n} \leqslant \psi_{u'',n},\ \forall u'' \in \mathcal{U}_n, \tag{37}$$

Algorithm 1 Iterative optimization for solving (31).

repeat
  for $n = 1 : N$ **do**
    for $u \in \mathcal{U}_n$ **do**
      Find the point in $\{0, \varepsilon, 2\varepsilon, \ldots, W_u\}$ that minimizes (39a) and satisfies (39b) and (39c) using one-dimensional search, and let it be $w^*_{u,n}$;
      Update $\boldsymbol{g}_{u,n}$ based on $w^*_{u,n}$ and (9);
      Update $\boldsymbol{f}_{u,n}, \forall u'' \in \mathcal{U}_n$ based on (26) and the new $\boldsymbol{g}_{u,n}$;
    end for
  end for
until Convergence;

Algorithm 2 Iterative optimization for solving (15).

Initialize $\boldsymbol{p}_{\mathcal{U}}$ and $\boldsymbol{w}_{\mathcal{U}}$;
Compute $\boldsymbol{F}_{\mathcal{U}}$ and $\boldsymbol{\mu}_{\mathcal{U}}$ based on (26) and Theorem 2;
repeat
  Update η based on Theorem 1;
  repeat
    Update $\boldsymbol{q}_{\mathcal{U}}$ by solving (30) and set $p_{u,n} = q^2_{u,n}, \forall u \in \mathcal{U}, n \in \mathcal{N}$;
    Update $\boldsymbol{w}_{\mathcal{U}}$ and $\boldsymbol{F}_{\mathcal{U}}$ using Algorithm 1;
    Update $\boldsymbol{\mu}_{\mathcal{U}}$ based on Theorem 2;
  until Convergence;
until Convergence;

where

$$\psi_{u'',n} = \frac{1}{\tau_n \mathrm{e}^{\mu_{u'',n}-1}} \left[\tau_n \mu_{u'',n} - \sum_{n' \in \mathcal{N}_{u''} \setminus \{n\}} \tau_{n'} \left(\mathrm{e}^{\mu_{u'',n'}-1} \mathrm{MSE}_{u'',n'} - \mu_{u'',n'} \right) - \hat{s} \right] - 1. \tag{38}$$

Based on (35) and (37), the problem (32) can be equivalently transformed to

$$\max_{w_{u,n}} \quad \sum_{u'' \in \mathcal{U}_n} \mathrm{e}^{\mu_{u'',n}-1} p_{u'',n} \boldsymbol{g}^{\mathrm{H}}_{u'',n} \left(p_{u,n} \boldsymbol{g}_{u,n} \boldsymbol{g}^{\mathrm{H}}_{u,n} + \boldsymbol{Q}_{u,n} \right)^{-1} \boldsymbol{g}_{u'',n} \tag{39a}$$

$$\text{s.t.} \quad -p_{u'',n} \boldsymbol{g}^{\mathrm{H}}_{u'',n} \left(p_{u,n} \boldsymbol{g}_{u,n} \boldsymbol{g}^{\mathrm{H}}_{u,n} + \boldsymbol{Q}_{u,n} \right)^{-1} \boldsymbol{g}_{u'',n} \leqslant \psi_{u'',n}, \ \forall u'' \in \mathcal{U}_n, \tag{39b}$$

$$0 \leqslant w_{u,n} \leqslant W_u. \tag{39c}$$

We divide the interval $[0, W_u]$ into $K+1$ equal-length subintervals, yielding $K+1$ boundary points $\{0, \varepsilon, 2\varepsilon, \ldots, W_u\}$, where $\varepsilon = W_u/K$ is the length of each subinterval. Then, a near-optimal solution of (39) can be easily found by using the exhaustive search (ES) (or one-dimensional search) method[3]. Once the near-optimal $w^*_{u,n}$ is obtain, we update $\boldsymbol{g}_{u,n}$ using (9), and then $\boldsymbol{f}_{u'',n}, \forall u'' \in \mathcal{U}_n$ based on (26). The iterative steps for solving (31) are summarized in Algorithm 1.

3.2.3 *Optimization of* $\boldsymbol{\mu}_{\mathcal{U}}$

Given $\boldsymbol{p}_{\mathcal{U}}$, $\boldsymbol{w}_{\mathcal{U}}$, and $\boldsymbol{F}_{\mathcal{U}}$, the optimal $\boldsymbol{\mu}^*_{\mathcal{U}}$ can be obtained based on Theorem 2. So far, we have shown how to optimize the variables in each iteration. The main steps for solving the original problem (15) are summarized in Algorithm 2.

3.3 Convergence and complexity analysis

In this subsection, we analyze the convergence and complexity of the proposed Algorithms 1 and 2.

3.3.1 *Convergence analysis*

As shown in Algorithm 1, for each time block n and user u in $\mathcal{U}_n$, a near-optimal $w_{u,n}$ can be obtained by solving (39). Then, $\boldsymbol{f}_{u,n}, \forall u'' \in \mathcal{U}_n$ are optimally updated based on (26). Therefore, in each iteration, the objective function of (31) decreases. Since the MSE is obviously limited, the convergence of Algorithm 1 is guaranteed. In Algorithm 2, η, $\boldsymbol{q}_{\mathcal{U}}$, $(\boldsymbol{w}_{\mathcal{U}}$, $\boldsymbol{F}_{\mathcal{U}})$, and $\boldsymbol{\mu}_{\mathcal{U}}$ are iteratively updated. In each step, the

3) Here we say that the solution is near-optimal since the variable $w_{u,n}$ is continuous and it is impossible to really perform ES. Therefore, we discretize the interval and carry out the one-dimensional search, whose performance depends on the size of the search step ε.

optimal or near-optimal solutions of these variables can be obtained, making the corresponding objective function decrease. Since the total energy consumption is limited, Algorithm 2 thus converges.

3.3.2 *Complexity analysis*

To assess the computational complexity of the proposed algorithms, we quantify the total number of floating-point operations (FLOPs), where one FLOP corresponds to either a complex multiplication or a complex summation. The complexity is expressed as a polynomial function of the dimensions of the involved matrices. For simplicity, we retain only the leading (i.e., highest-order or dominant) terms in the polynomial, as these terms primarily determine the asymptotic behavior of the complexity.

The complexity of Algorithm 1 primarily arises from the channel inversions performed in step 4, which involves a one-dimensional search to identify the near-optimal antenna position $w_{u,n}$. In this step, the objective function and constraints of (39) have to be computed $K+1$ times, with each computation requiring $|\mathcal{U}_n|$ channel inversions. Given that the inversion of an $M \times M$ Hermitian matrix requires $\mathcal{O}(M^3)$ FLOPs, each inner iteration of Algorithm 1 incurs a complexity of $\mathcal{O}(M^3|\mathcal{U}_n|(K+1))$. Algorithm 1 consists of N outer iterations, with each outer iteration involving $|\mathcal{U}_n|$ inner iterations. Since $|\mathcal{U}_n|$ varies with n and satisfies $|\mathcal{U}_n| \leqslant |\mathcal{U}| = U$, we simplify the complexity analysis by replacing $|\mathcal{U}_n|$ with U. Then, the overall complexity of Algorithm 1 is $\mathcal{O}(M^3U^2N(K+1))$. In Algorithm 2, the main complexity lies in step 7, i.e., using Algorithm 1 to update $\boldsymbol{w}_{\mathcal{U}}$ and $\boldsymbol{F}_{\mathcal{U}}$. Let ν_1 and ν_2 denote the numbers of outer and inner iterations of Algorithm 2. Then, Algorithm 2 involves a total complexity of $\mathcal{O}(\nu_1\nu_2M^3U^2N(K+1))$.

4 Simulation results

In this section, we evaluate the performance of the considered system and also the proposed algorithms by simulation. We consider the UMa (Urban Macro) environment. According to [49], the distance-dependent pathloss of the l-th path of user u is $\mathrm{PL}_{u,l} = 32.4+20\log_{10} f_c+30\log_{10} \varDelta_u+\varsigma_{u,l}$, where $\varDelta_u$ is the distance in meters between user u and the BS, and $\varsigma_{u,l} \sim \mathcal{N}(0,\sigma_{\mathrm{SF}}^2)$ is the log-normal shadow fading. σ_{SF} is the standard deviation of the shadow fading and is set to be 7.8 dB here. The mmWave system operates at a carrier frequency $f_c = 28$ GHz and the bandwidth is $B = 20$ MHz. The noise power is $\sigma^2 = -94$ dBm. In addition, we set $J = 10$, $\delta = 1/10$, $\gamma = 1$, $\kappa = 10^{-29}$, $\xi = 1/10$, $\epsilon_0 = 10^{-3}$, and $\rho = 1/2$. The parameter C_u is uniformly distributed in $[10^4, 3\times 10^4]$ cycles/sample. Each user has $D_u = 500$ data samples, which are randomly selected from the dataset with equal probability. The computation capacity of each user is $G_u = 1$ GHz. For convenience, we assume equal FAS size and number of paths for all users, i.e., $W_u = W,\ L_u = L, \forall u \in \mathcal{U}$. The other parameters are specified in each figure. We compare the performance of FAS with the FPA strategy, which follows the same steps as Algorithm 2, but omits updates to $\boldsymbol{w}_{\mathcal{U}}$. All results are obtained by averaging over 1000 independent channel realizations, with each realization generated randomly according to the specified distribution.

Figure 3 depicts the convergence behaviors of the proposed Algorithm 2 under different configurations of U. It can be seen from this figure that the total energy consumption decreases monotonically during the iterative procedure and converges rapidly in about 4 iterations for all configurations. Moreover, Figure 3 shows that the energy consumption grows with U. This is consistent with intuition since more users result in more computation load, transmission data, and co-channel interference.

Figure 4 investigates the effect of the number of BS antennas and the transmit data size of the users. Several observations can be made from this figure. First, the total energy consumption decreases with M. This is because when the BS has more antennas, it becomes more powerful in receiving the desired signal and compressing the co-channel interference. Second, it can be seen that when M is small, in contrast to the FPA scheme, FAS can dramatically decrease the energy consumption. For example, when $M = 2$ and $s = 60$ kbits. FPA requires about 660 J to complete the task, while FAS requires only 420 J. However, as the number of antennas increases, this advantage becomes less pronounced. This suggests that FAS and the multi-antenna techniques can be complementary. In addition, we see that as expected, the energy consumption increases with the transmit data size s, and when FAS demonstrates a more significant advantage over FPA in terms of the energy efficiency.

Figure 5 compares the performance of FAS with that of FPA under different configurations of L and U. Besides the observation that the energy consumption increases with U, which has already been verified by Figure 3, Figure 5 also shows that the energy consumption decreases with L, indicating that a rich scattering environment enhances communication effectiveness. In addition, it can be seen that the gap

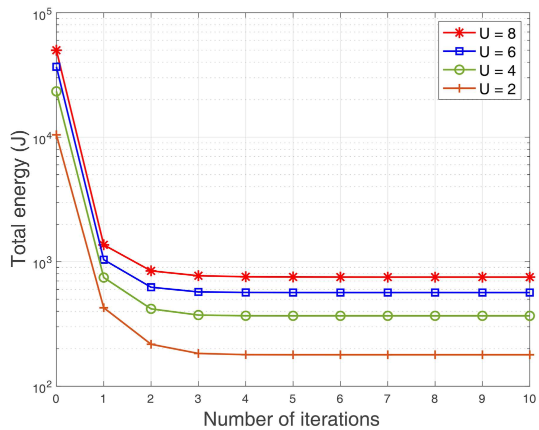

Figure 3 (Color online) Convergence behaviors of Algorithm 2 with $W = 10\lambda$, $M = 4$, $L = 5$, and $s = 60$ kbits.

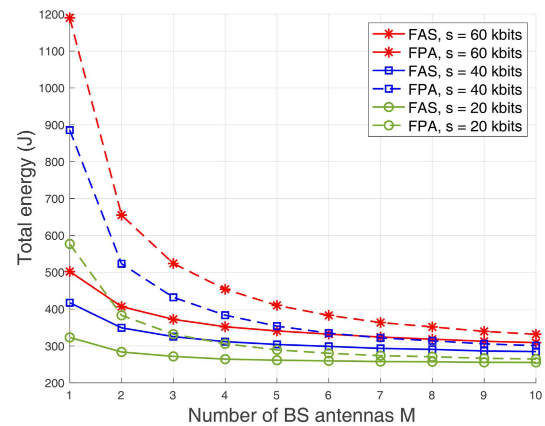

Figure 4 (Color online) Total energy consumption versus the number of BS antennas with $U = 4$, $W = 10\lambda$, and $L = 5$.

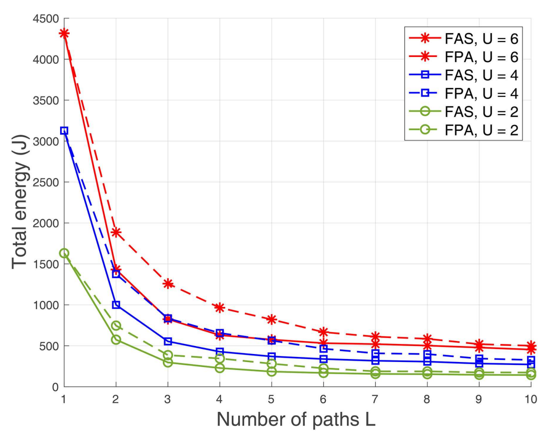

Figure 5 (Color online) Total energy consumption versus the number of paths with $W = 10\lambda$, $M = 4$, and $s = 60$ kbits.

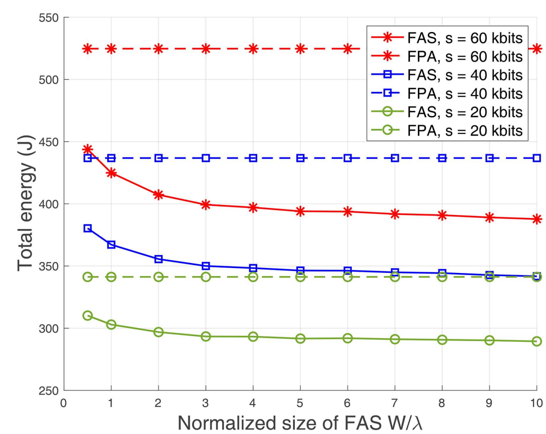

Figure 6 (Color online) Total energy consumption versus the normalized size of FAS with $U = 4$, $M = 4$, and $L = 5$.

between FAS and FPA first increases and then decreases with L. Interestingly, when $L = 1$, i.e., there is only one path between each user and the BS, the energy consumption required by the methods using FAS and FPA is exactly the same. Therefore, using FAS in this case does not benefit the performance. Note that similar observations have also been made in the sum capacity maximization problem of a multi-user uplink communication system [11], and have been theoretically analyzed (see [11, Lemma 1]).

Figure 6 investigates the effect of the FAS size W. Note that W does not affect the performance of the FPA method. Therefore, for any given s, the total energy consumed by the FPA method is a constant. It is seen from this figure that as W increases, the total energy consumed by the FAS strategy decreases significantly at the beginning and then saturates. This is because we use the same $K = 100$ for all configurations when solving the problem (39) in Algorithm 1. Then, when W is relatively small, increasing W can significantly reduce the correlation among the $K + 1$ sampling points. However, when W becomes large, the spatial correlation among the $K+1$ sampling points is no longer significant. In this case, for a fixed K, further increases in W will not result in a significant reduction in energy consumption.

5 Conclusion

This paper investigated FAS-assisted FL in a multi-user communication system, aiming to improve the energy efficiency of the FL process. We addressed this by formulating an optimization problem that minimizes the total energy consumed for user-side computation and data transmission, subject to latency and data size constraints. Due to the coupling of variables such as learning accuracy, transmit power, antenna positions, and BS receivers, the considered problem is inherently non-convex and challenging to solve. To overcome this, we proposed a low-complexity algorithm that works iteratively and provides a near-optimal solution. Simulation results have confirmed the algorithm's effectiveness and highlighted the advantages of FAS over FPA.

Acknowledgements This work of Hao XU was supported by Fundamental Research Funds for the Central Universities (Grant No. 2242025R10001). This work of Kai-Kit WONG, Wee Kiat NEW, and Farshad Rostami GHADI was supported by Engineering and Physical Sciences Research Council (EPSRC) (Grant No. EP/W026813/1). This work of Chongwen HUANG was supported by China National Key R&D Program (Grant Nos. 2021YFA1000500, 2023YFB2904804), National Natural Science Foundation of China (Grant Nos. 62331023, 62394292), Zhejiang Provincial Science and Technology Plan Project (Grant No. 2024C01033), and Zhejiang University Global Partnership Fund. This work of Chao WANG was supported by National Natural Science Foundation of China (Grant No. 62371357) and Funds for International Cooperation and Exchange of National Natural Science Foundation of China (Grant No. 12411530120).

References

1 Wong K, Tong K, Zhang Y, et al. Fluid antenna system for 6G: when Bruce Lee inspires wireless communications. Electron Lett, 2020, 56: 1288–1290
2 Wong K K, Shojaeifard A, Tong K F, et al. Fluid antenna systems. IEEE Trans Wireless Commun, 2020, 20: 1950–1962
3 Wong K K, New W K, Hao X, et al. Fluid antenna system—part I: preliminaries. IEEE Commun Lett, 2023, 27: 1919–1923
4 New W K, Wong K K, Xu H, et al. A tutorial on fluid antenna system for 6G networks: encompassing communication theory, optimization methods and hardware designs. IEEE Commun Surv Tut, 2024. doi: 10.1109/COMST.2024.3498855
5 Wu T, Zhi K, Yao J, et al. Fluid antenna systems enabling 6G: principles, applications, and research directions. 2024. ArXiv:2412.03839
6 New W K, Wong K K, Xu H, et al. Fluid antenna system: new insights on outage probability and diversity gain. IEEE Trans Wireless Commun, 2024, 23: 128–140
7 Ghadi F R, Wong K K, López-Martínez F J, et al. Copula-based performance analysis for fluid antenna systems under arbitrary fading channels. IEEE Commun Lett, 2023, 27: 3068–3072
8 Ghadi F R, Wong K K, López-Martínez F J, et al. A Gaussian copula approach to the performance analysis of fluid antenna systems. IEEE Trans Wireless Commun, 2024, 23: 17573–17585
9 New W K, Wong K K, Xu H, et al. An information-theoretic characterization of MIMO-FAS: optimization, diversity-multiplexing tradeoff and q-outage capacity. IEEE Trans Wireless Commun, 2024, 23: 5541–5556
10 Mei W, Wei X, Ning B, et al. Movable-antenna position optimization: a graph-based approach. IEEE Wireless Commun Lett, 2024, 13: 1853–1857
11 Xu H, Wong K K, New W K, et al. Capacity maximization for FAS-assisted multiple access channels. 2023. ArXiv:2311.11037
12 Shao X, Jiang Q, Zhang R. 6D movable antenna based on user distribution: modeling and optimization. 2024. ArXiv:2403.08123
13 Shao X, Zhang R, Jiang Q, et al. 6D movable antenna enhanced wireless network via discrete position and rotation optimization. 2024. ArXiv:2403.17122
14 Xu H, Wong K K, New W K, et al. Coding-enhanced cooperative jamming for secret communication in fluid antenna systems. IEEE Commun Lett, 2024, 28: 1991–1995
15 Ghadi F R, Wong K K, Javier López-Martínez F, et al. Physical layer security over fluid antenna systems: secrecy performance analysis. IEEE Trans Wireless Commun, 2024, 23: 18201–18213
16 Wong K K, Tong K F. Fluid antenna multiple access. IEEE Trans Wireless Commun, 2021, 21: 4801–4815
17 Xu H, Wong K K, New W K, et al. Revisiting outage probability analysis for two-user fluid antenna multiple access system. IEEE Trans Wireless Commun, 2024, 23: 9534–9548
18 Ghadi F R, Wong K K, Kaveh M, et al. Fluid antenna multiple access with simultaneous non-unique decoding in strong interference channel. 2024. ArXiv:2410.20930
19 Xu H, Zhou G, Wong K K, et al. Channel estimation for FAS-assisted multiuser mmWave systems. IEEE Commun Lett, 2023, 28: 632–636
20 New W K, Wong K K, Xu H, et al. Channel estimation and reconstruction in fluid antenna system: oversampling is essential. IEEE Trans Wireless Commun, 2025, 24: 309–322
21 Wong K K, Chae C B, Tong K F. Compact ultra massive antenna array: a simple open-loop massive connectivity scheme. IEEE Trans Wireless Commun, 2024, 23: 6279–6294
22 New W K, Wong K K, Xu H, et al. Fluid antenna system enhancing orthogonal and non-orthogonal multiple access. IEEE Commun Lett, 2024, 28: 218–222
23 Zheng J, Wu T, Lai X, et al. FAS-assisted NOMA short-packet communication systems. IEEE Trans Veh Technol, 2024, 73: 10732–10737
24 Wang C, Li Z, Wong K K, et al. AI-empowered fluid antenna systems: opportunities, challenges, and future directions. IEEE Wireless Commun, 2024, 31: 34–41
25 Wang C, Li G, Zhang H, et al. Fluid antenna system liberating multiuser MIMO for ISAC via deep reinforcement learning. IEEE Trans Wireless Commun, 2024, 23: 10879–10894
26 Waqar N, Wong K K, Chae C B, et al. Opportunistic fluid antenna multiple access via team-inspired reinforcement learning. IEEE Trans Wireless Commun, 2024, 23: 12068–12083
27 Zou J, Xu H, Wang C, et al. Shifting the ISAC trade-off with fluid antenna systems. IEEE Wireless Commun Lett, 2024, 13: 3479–3483
28 Zhou L, Yao J, Jin M, et al. Fluid antenna-assisted ISAC systems. IEEE Wireless Commun Lett, 2024, 13: 3533–3537
29 Lu S, Liu F, Li Y, et al. Integrated sensing and communications: recent advances and ten open challenges. IEEE Int Things J, 2024, 11: 19094–19120
30 Ghadi F R, Wong K K, New W K, et al. On performance of RIS-aided fluid antenna systems. IEEE Wireless Commun Lett, 2024, 13: 2175–2179
31 Hu G, Wu Q, Xu D, et al. Intelligent reflecting surface-aided wireless communication with movable elements. IEEE Wireless Commun Lett, 2024, 13: 1173–1177
32 Xu W, Huang Y M, Wang W, et al. Toward ubiquitous and intelligent 6G networks: from architecture to technology. Sci China Inf Sci, 2023, 66: 130300
33 Sun Y, Peng M, Zhou Y, et al. Application of machine learning in wireless networks: key techniques and open issues. IEEE Commun Surv Tut, 2019, 21: 3072–3108
34 Xu J D, Yuen C, Huang C W, et al. Reconfiguring wireless environments via intelligent surfaces for 6G: reflection, modulation, and security. Sci China Inf Sci, 2023, 66: 130304
35 Wang S, Tuor T, Salonidis T, et al. When edge meets learning: adaptive control for resource-constrained distributed machine learning. In: Proceedings of IEEE Conference on Computer Communications, Honolulu, 2018. 63–71

36 Xu W, Yang Z, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. IEEE J Sel Top Signal Process, 2023, 17: 9–39
37 Xu W, Wu J, Jin S, et al. Disentangled representation learning empowered CSI feedback using implicit channel reciprocity in FDD massive MIMO. IEEE Trans Wireless Commun, 2024, 23: 15169–15184
38 Chen M, Gunduz D, Huang K, et al. Distributed learning in wireless networks: recent progress and future challenges. IEEE J Sel Areas Commun, 2021, 39: 3579–3605
39 Yang Z, Chen M, Wong K K, et al. Federated learning for 6G: applications, challenges, and opportunities. Engineering, 2022, 8: 33–41
40 Chen M, Poor H V, Saad W, et al. Convergence time optimization for federated learning over wireless networks. IEEE Trans Wireless Commun, 2020, 20: 2457–2471
41 Chen M, Yang Z, Saad W, et al. A joint learning and communications framework for federated learning over wireless networks. IEEE Trans Wireless Commun, 2020, 20: 269–283
42 Tran N H, Bao W, Zomaya A, et al. Federated learning over wireless networks: optimization model design and analysis. In: Proceedings of IEEE Conference on Computer Communications, Paris, 2019. 1387–1395
43 Yang Z, Chen M, Saad W, et al. Energy efficient federated learning over wireless communication networks. IEEE Trans Wireless Commun, 2020, 20: 1935–1949
44 Zhang J, Rao J, Li Z, et al. A novel pixel-based reconfigurable antenna applied in fluid antenna systems with high switching speed. IEEE Open J Antennas Propag, 2025, 6: 212–228
45 Mao Y, Zhang J, Letaief K B. Dynamic computation offloading for mobile-edge computing with energy harvesting devices. IEEE J Sel Areas Commun, 2016, 34: 3590–3605
46 Akdeniz M R, Liu Y, Samimi M K, et al. Millimeter wave channel modeling and cellular capacity evaluation. IEEE J Sel Areas Commun, 2014, 32: 1164–1179
47 Xu H, Pan C, Xu W, et al. Robust beamforming with pilot reuse scheduling in a heterogeneous cloud radio access network. IEEE Trans Veh Technol, 2018, 67: 7242–7256
48 Xu H, Stuber G L, Xu W, et al. Robust transmission design for multicell D2D underlaid cellular networks. IEEE Trans Veh Technol, 2018, 67: 5922–5936
49 3GPP. Technical Specification Group Radio Access Network; Study on Channel Model for Frequencies from 0.5 to 100 GHZ. TR 38.901 v16.1.0, 2019

Appendix A Proof of Theorem 1

According to (18), the first-order derivative of E w.r.t. η is

$$\frac{\partial E}{\partial \eta} = \frac{-\alpha_1(\frac{1}{\eta} + \ln\eta) + \alpha_1 + \alpha_2}{(1-\eta)^2}. \tag{A1}$$

For convenience, denote function $r(\eta) = \frac{1}{\eta} + \ln\eta$. Its first-order derivative is given by

$$\frac{\partial r}{\partial \eta} = \frac{\eta - 1}{\eta^2}, \tag{A2}$$

which is non-positive in $(0, 1]$ and non-negative in $[1, +\infty)$. Therefore, when η increases from 0 to 1, $r(\eta)$ decreases monotonically from $+\infty$ to 1. Since α_1 and α_2 are both positive, the numerator of (A1) increases monotonically from $-\infty$ to α_2 as η increases from 0 to 1. Then, there must exist a unique zero point of the numerator of (A1) in the interval $(0, 1]$. Let $\eta_0 \in (0, 1]$ denote this zero point, i.e., it makes the following equation holds:

$$\frac{1}{\eta} + \ln\eta = 1 + \frac{\alpha_2}{\alpha_1}. \tag{A3}$$

Since $\frac{1}{\eta} + \ln\eta$ decreases monotonically w.r.t. η in $(0, 1]$, η_0 can be easily found using the bisection method.

Accordingly, we know that $\partial E/\partial\eta$ is non-positive in $(0, \eta_0]$ and non-negative in $[\eta_0, +\infty)$, indicating that E decreases w.r.t. η in $(0, \eta_0]$ and then increases in $[\eta_0, +\infty)$. Since $\eta_0 \in (0, 1]$ and η belongs to $[\alpha_0, 1]$ (see constraint (22b)), we know that if $\alpha_0 < \eta_0$, the optimal solution of (22) is η_0. Otherwise, it is α_0. Theorem 1 is thus proven.

Appendix B Proof of Theorem 2

Let $(\boldsymbol{p}_{\mathcal{U}}^{\circledast}, \boldsymbol{w}_{\mathcal{U}}^{\circledast}, \boldsymbol{F}_{\mathcal{U}}^{\circledast})$ denote the optimal solution of (27) and $\mathrm{MSE}_{u,n}^{\circledast}$ represent the corresponding MSE of user u at time block n. Furthermore, let $\mu_{u,n}^{\circledast} = 1 - \ln \mathrm{MSE}_{u,n}^{\circledast}$. Substituting $\mu_{u,n}^{\circledast}$ and $\mathrm{MSE}_{u,n}^{\circledast}$ into (28b), we have

$$\begin{aligned}\sum_{n\in\mathcal{N}_u} \tau_n \left(\mathrm{e}^{\mu_{u,n}^{\circledast}-1}\mathrm{MSE}_{u,n}^{\circledast} - \mu_{u,n}^{\circledast}\right) &= \sum_{n\in\mathcal{N}_u} \tau_n \ln \mathrm{MSE}_{u,n}^{\circledast} \\ &\leqslant -\hat{s},\ \forall u \in \mathcal{U},\end{aligned} \tag{B1}$$

where the last step follows from the fact that $\mathrm{MSE}_{u,n}^{\circledast}$ makes (27b) hold. Then, we know that $(\boldsymbol{p}_{\mathcal{U}}^{\circledast}, \boldsymbol{w}_{\mathcal{U}}^{\circledast}, \boldsymbol{F}_{\mathcal{U}}^{\circledast}, \boldsymbol{\mu}_{\mathcal{U}}^{\circledast})$ can make all the constraints of (28) satisfied and is thus a solution of (28).

Let $(\boldsymbol{p}_{\mathcal{U}}^{*}, \boldsymbol{w}_{\mathcal{U}}^{*}, \boldsymbol{F}_{\mathcal{U}}^{*}, \boldsymbol{\mu}_{\mathcal{U}}^{*})$ denote the optimal solution of (28) and $\mathrm{MSE}_{u,n}^{*}$ represent the corresponding MSE of user u at time block n. In the following we first show that (29) is true, based on which we then know that $(\boldsymbol{p}_{\mathcal{U}}^{*}, \boldsymbol{w}_{\mathcal{U}}^{*}, \boldsymbol{F}_{\mathcal{U}}^{*})$ is a solution of (27). Based on (25), the constraint (28b) can be rewritten as

$$\sum_{n\in\mathcal{N}_u} \tau_n \left(\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}_{u,n} - \mu_{u,n}\right)$$

$$= \sum_{n\in\mathcal{N}_u} \tau_n \left\{ \mathrm{e}^{\mu^*_{u,n}-1} \left[\left(\sqrt{p^*_{u,n}} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u\}} p^*_{u',n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u',n} \boldsymbol{g}^{*\mathrm{H}}_{u',n} \boldsymbol{f}^*_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u,n} \right] - \mu^*_{u,n} \right\}$$
$$\leqslant -\hat{s},\ \forall u\in\mathcal{U}. \tag{B2}$$

For convenience, we define $q^*_{u,n} = \sqrt{p^*_{u,n}}$ and rewrite (B2) equivalently as follows:

$$\sum_{n\in\mathcal{N}_u} \tau_n \left\{ \mathrm{e}^{\mu^*_{u,n}-1} \left[\left(q^*_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u\}} q^{*2}_{u',n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u',n} \boldsymbol{g}^{*\mathrm{H}}_{u',n} \boldsymbol{f}^*_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u,n} \right] - \mu^*_{u,n} \right\}$$
$$\leqslant -\hat{s},\ \forall u\in\mathcal{U}. \tag{B3}$$

To prove (29), we first give Lemma B1, which provides an upper bound on the optimal transmit power.

Lemma B1. In the optimal case of problem (28), the transmit power satisfies

$$\sqrt{p^*_{u,n}} = q^*_{u,n} \leqslant \frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n}},\ \forall u\in\mathcal{U},\ n\in\mathcal{N}_u. \tag{B4}$$

Proof. See Appendix C.

Now we prove (29) by contraposition. Assume that Eq. (29) is not true. Then, there must exist a user $u\in\mathcal{U}$ and a time block $n\in\mathcal{N}_u$ such that

$$\mu^*_{u,n} \neq 1 - \ln \mathrm{MSE}^*_{u,n}. \tag{B5}$$

For this given user $u\in\mathcal{U}$, the corresponding inequality in (B3) can be rewritten as

$$\tau_n \left\{ \mathrm{e}^{\mu^*_{u,n}-1} \left[\left(q^*_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u\}} q^{*2}_{u',n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u',n} \boldsymbol{g}^{*\mathrm{H}}_{u',n} \boldsymbol{f}^*_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u,n} \right] - \mu^*_{u,n} \right\}$$
$$+ \sum_{n'\in\mathcal{N}_u\setminus\{n\}} \tau_{n'} \left(\mathrm{e}^{\mu^*_{u,n'}-1} \mathrm{MSE}^*_{u,n'} - \mu^*_{u,n'} \right) \leqslant -\hat{s}, \tag{B6}$$

in which we separate the summation terms at time block n from those at time blocks $n'\in\mathcal{N}_u\setminus\{n\}$ for convenience. Note that since $\mathrm{MSE}^*_{u,n}\geqslant 0$, $\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}^*_{u,n} - \mu_{u,n}$ is a convex function of $\mu_{u,n}$ and its unique optimal solution is

$$\mu^\star_{u,n} = \arg\min_{\mu_{u,n}} \left(\mathrm{e}^{\mu_{u,n}-1} \mathrm{MSE}^*_{u,n} - \mu_{u,n} \right)$$
$$= 1 - \ln \mathrm{MSE}^*_{u,n}. \tag{B7}$$

Then, by replacing $\mu^*_{u,n}$ in (B6) with $\mu^\star_{u,n}$, we have

$$\tau_n \left\{ \mathrm{e}^{\mu^\star_{u,n}-1} \left[\left(q^*_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u\}} q^{*2}_{u',n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u',n} \boldsymbol{g}^{*\mathrm{H}}_{u',n} \boldsymbol{f}^*_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u,n} \right] - \mu^\star_{u,n} \right\}$$
$$+ \sum_{n'\in\mathcal{N}_u\setminus\{n\}} \tau_{n'} \left(\mathrm{e}^{\mu^*_{u,n'}-1} \mathrm{MSE}^*_{u,n'} - \mu^*_{u,n'} \right) < -\hat{s}. \tag{B8}$$

Note that the inequality in (B8) strictly holds, while in (B6), the inequality may hold with equality. This is because $\mu^\star_{u,n}$ is the unique point that minimizes $\mathrm{e}^{\mu_{u,n}-1}\mathrm{MSE}^*_{u,n} - \mu_{u,n}$ (see (B7)), while $\mu^*_{u,n}$ cannot due to (B5). As discussed in Appendix C, the term $\tau_n \mathrm{e}^{\mu^\star_{u,n}-1} \left(q^*_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2$ in (B8) is a quadratic function of $q^*_{u,n}$ with an upward opening and its zero point is $\frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n}}$. Since $q^*_{u,n} \leqslant \frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n}}$ (see Lemma B1), we can always find a new $q^\star_{u,n}$ such that

$$q^\star_{u,n} < q^*_{u,n}, \tag{B9}$$

and

$$\tau_n \left\{ \mathrm{e}^{\mu^\star_{u,n}-1} \left[\left(q^\star_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u,n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u\}} q^{*2}_{u',n} \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{g}^*_{u',n} \boldsymbol{g}^{*\mathrm{H}}_{u',n} \boldsymbol{f}^*_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u,n} \right] - \mu^\star_{u,n} \right\}$$
$$+ \sum_{n'\in\mathcal{N}_u\setminus\{n\}} \tau_{n'} \left(\mathrm{e}^{\mu^*_{u,n'}-1} \mathrm{MSE}^*_{u,n'} - \mu^*_{u,n'} \right) = -\hat{s}. \tag{B10}$$

Note that as discussed in Appendix C, for given $u\in\mathcal{U}$ and $n\in\mathcal{N}_u$, besides the inequality in (B3) with u, which is exactly (B8), $q^*_{u,n}$ also exists in $|\mathcal{U}_n|-1$ other inequalities in (B3) with $u''\in\mathcal{U}_n\setminus\{u\}$ (see (C2)). In particular, for each $u''\in\mathcal{U}_n\setminus\{u\}$, $q^*_{u,n}$ exists in the term $\tau_n \mathrm{e}^{\mu^*_{u'',n}-1} q^{*2}_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u'',n} \boldsymbol{g}^*_{u,n} \boldsymbol{g}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u'',n}$. Due to (B9), it is obvious that

$$\tau_n \mathrm{e}^{\mu^*_{u'',n}-1} q^{\star 2}_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u'',n} \boldsymbol{g}^*_{u,n} \boldsymbol{g}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u'',n} < \tau_n \mathrm{e}^{\mu^*_{u'',n}-1} q^{*2}_{u,n} \boldsymbol{f}^{*\mathrm{H}}_{u'',n} \boldsymbol{g}^*_{u,n} \boldsymbol{g}^{*\mathrm{H}}_{u,n} \boldsymbol{f}^*_{u'',n},\ \forall u''\in\mathcal{U}_n\setminus\{u\}. \tag{B11}$$

Therefore, we know that by replacing $p^{*}_{u,n}$ and $\mu^{*}_{u,n}$ in $(\boldsymbol{p}^{*}_{\mathcal{U}}, \boldsymbol{w}^{*}_{\mathcal{U}}, \boldsymbol{F}^{*}_{\mathcal{U}}, \boldsymbol{\mu}^{*}_{\mathcal{U}})$ with $p^{\star}_{u,n} = q^{\star 2}_{u,n}$ and $\mu^{\star}_{u,n}$, we can obtain a new feasible solution for (28), and due to (B9), the objective function value can be further decreased. This contradicts the assumption that $(\boldsymbol{p}^{*}_{\mathcal{U}}, \boldsymbol{w}^{*}_{\mathcal{U}}, \boldsymbol{F}^{*}_{\mathcal{U}}, \boldsymbol{\mu}^{*}_{\mathcal{U}})$ is the optimal solution of (28). Therefore, Eq. (29) is true.

On the other hand, due to (29), we have

$$\sum_{n\in\mathcal{N}_u} \tau_n \left(\mathrm{e}^{\mu^{*}_{u,n}-1}\mathrm{MSE}^{*}_{u,n} - \mu^{*}_{u,n}\right) = \sum_{n\in\mathcal{N}_u} \tau_n \ln \mathrm{MSE}^{*}_{u,n} \leqslant -\hat{s},\ \forall u \in \mathcal{U}. \tag{B12}$$

Therefore, $(\boldsymbol{p}^{*}_{\mathcal{U}}, \boldsymbol{w}^{*}_{\mathcal{U}}, \boldsymbol{F}^{*}_{\mathcal{U}})$ is also a solution of (28). Obviously, the optimal objective function value of (27) resulted from $(\boldsymbol{p}^{\circledast}_{\mathcal{U}}, \boldsymbol{w}^{\circledast}_{\mathcal{U}}, \boldsymbol{F}^{\circledast}_{\mathcal{U}})$ is equal to that of (28) resulted from $(\boldsymbol{p}^{*}_{\mathcal{U}}, \boldsymbol{w}^{*}_{\mathcal{U}}, \boldsymbol{F}^{*}_{\mathcal{U}}, \boldsymbol{\mu}^{*}_{\mathcal{U}})$, since otherwise the objective function value can be further decreased by obtaining a new solution for (27) from $(\boldsymbol{p}^{*}_{\mathcal{U}}, \boldsymbol{w}^{*}_{\mathcal{U}}, \boldsymbol{F}^{*}_{\mathcal{U}}, \boldsymbol{\mu}^{*}_{\mathcal{U}})$ or obtaining a new solution for (28) from $(\boldsymbol{p}^{\circledast}_{\mathcal{U}}, \boldsymbol{w}^{\circledast}_{\mathcal{U}}, \boldsymbol{F}^{\circledast}_{\mathcal{U}})$. Accordingly, we know that Eqs. (27) and (28) are equivalent in terms of that they have the optimal solution. Theorem 2 is thus proven.

Appendix C Proof of Lemma B1

For any given time block $n \in \mathcal{N}$ and user $u \in \mathcal{U}_n$, from (B3) we know that only inequalities with $u \in \mathcal{U}_n$ contain $q^{*}_{u,n}$. This is because, during the time block n, only users in $\mathcal{U}_n$ are transmitting their signals. Therefore, the signal of user u only affects the communication of users in $\mathcal{U}_n$. Then, $q^{*}_{u,n}$ appears in $|\mathcal{U}_n|$ inequalities in (B3), i.e., the $\tau_n \mathrm{e}^{\mu^{*}_{u,n}-1}\left(q^{*}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n} - 1\right)^2$ term in

$$\begin{aligned}
&\tau_n \mathrm{e}^{\mu^{*}_{u,n}-1}\left(q^{*}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n} - 1\right)^2 + \tau_n \left\{\mathrm{e}^{\mu^{*}_{u,n}-1}\left[\sum_{u'\in\mathcal{U}_n\setminus\{u\}} q^{*2}_{u',n}\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u',n}\boldsymbol{g}^{*\mathrm{H}}_{u',n}\boldsymbol{f}^{*}_{u,n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{f}^{*}_{u,n}\right] - \mu^{*}_{u,n}\right\}\\
&+ \sum_{n'\in\mathcal{N}_u\setminus\{n\}} \tau_{n'}\left(\mathrm{e}^{\mu^{*}_{u,n'}-1}\mathrm{MSE}^{*}_{u,n'} - \mu^{*}_{u,n'}\right) \leqslant -\hat{s},
\end{aligned} \tag{C1}$$

and the $\tau_n \mathrm{e}^{\mu^{*}_{u'',n}-1} q^{*2}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{g}^{*}_{u,n}\boldsymbol{g}^{*\mathrm{H}}_{u,n}\boldsymbol{f}^{*}_{u'',n}$ term in

$$\begin{aligned}
&\tau_n \mathrm{e}^{\mu^{*}_{u'',n}-1} q^{*2}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{g}^{*}_{u,n}\boldsymbol{g}^{*\mathrm{H}}_{u,n}\boldsymbol{f}^{*}_{u'',n}\\
&+ \tau_n \left\{\mathrm{e}^{\mu^{*}_{u'',n}-1}\left[\left(q^{*}_{u'',n}\boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{g}^{*}_{u'',n} - 1\right)^2 + \sum_{u'\in\mathcal{U}_n\setminus\{u'',u\}} q^{*2}_{u',n}\boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{g}^{*}_{u',n}\boldsymbol{g}^{*\mathrm{H}}_{u',n}\boldsymbol{f}^{*}_{u'',n} + \sigma^2 \boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{f}^{*}_{u'',n}\right] - \mu^{*}_{u'',n}\right\}\\
&+ \sum_{n'\in\mathcal{N}_{u''}\setminus\{n\}} \tau_{n'}\left(\mathrm{e}^{\mu^{*}_{u'',n'}-1}\mathrm{MSE}^{*}_{u'',n'} - \mu^{*}_{u'',n'}\right) \leqslant -\hat{s},\ \forall u'' \in \mathcal{U}_n \setminus \{u\}.
\end{aligned} \tag{C2}$$

From (26) we know that $\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n}$ is real and positive. The zero point of

$$\tau_n \mathrm{e}^{\mu^{*}_{u,n}-1}\left(q^{*}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n} - 1\right)^2 = 0 \tag{C3}$$

is $\frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n}}$. Since $\tau_n \mathrm{e}^{\mu^{*}_{u,n}-1} > 0$, $\tau_n \mathrm{e}^{\mu^{*}_{u,n}-1}\left(q^{*}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n} - 1\right)^2$ is a quadratic function of $q^{*}_{u,n}$ with an upward opening, and its value first decreases w.r.t. $q^{*}_{u,n}$ in $[0, \frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n}}]$ and then increases in $[\frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n}}, +\infty)$. Note that the $\tau_n \mathrm{e}^{\mu^{*}_{u'',n}-1} q^{*2}_{u,n}\boldsymbol{f}^{*\mathrm{H}}_{u'',n}\boldsymbol{g}^{*}_{u,n}\boldsymbol{g}^{*\mathrm{H}}_{u,n}\boldsymbol{f}^{*}_{u'',n}$ term in (C2) increases w.r.t. $q^{*}_{u,n}$ in $[0, +\infty]$. Therefore, if $q^{*}_{u,n} > \frac{1}{\boldsymbol{f}^{*\mathrm{H}}_{u,n}\boldsymbol{g}^{*}_{u,n}}$, we can always decrease $q^{*}_{u,n}$ such that Eqs. (C1) and (C2) still hold. Then, with the newly obtained $q^{*}_{u,n}$, all constraints of (28) hold, and its objective function is further decreased. This contradicts the assumption that $q^{*}_{u,n}$ is the optimal solution. Therefore, in the optimal case, Eq. (B4) is true. This completes the proof of Lemma B1.