Massive MIMO for Serving Federated Learning and Non-Federated Learning Users

Muhammad Farooq, *Member, IEEE*, Tung Thanh Vu, *Member, IEEE*, Hien Quoc Ngo, *Senior Member, IEEE*, and Le-Nam Tran, *Senior Member, IEEE*

***Abstract*— With its privacy preservation and communication efficiency, federated learning (FL) has emerged as a promising learning framework for beyond 5G wireless networks. It is anticipated that future wireless networks will jointly serve both FL and downlink non-FL user groups in the same time-frequency resource. While in the downlink of each FL iteration, both groups simultaneously receive data from the base station in the same time-frequency resource, the uplink of each FL iteration requires bidirectional communication to support uplink transmission for FL users and downlink transmission for non-FL users. To overcome this challenge, we present half-duplex (HD) and full-duplex (FD) communication schemes to serve both groups. More specifically, we adopt the massive multiple-input multiple-output technology and aim to maximize the minimum effective rate of non-FL users under a quality of service (QoS) latency constraint for FL users. Since the formulated problem is nonconvex, we propose a power control algorithm based on successive convex approximation to find a stationary solution. Numerical results show that the proposed solutions perform significantly better than the considered baselines schemes. Moreover, the FD-based scheme outperforms the HD-based counterpart in scenarios where the self-interference is small or moderate and/or the size of FL model updates is large.**

***Index Terms*— Massive multiple-input multiple-output (MIMO), federated learning (FL), resource allocation, successive convex approximation (SCA).**

I. Introduction

THE use of mobile phones and wearable devices enables continuous collection and transfer of data [2], [3], which has been the main driving force behind the explosive increase in data mobile traffic in recent years. Also, due to a constant growing interest in new features and tools, the computational power of these devices is increasing day by day. Thus, in many applications, part of data processing is carried out at user equipment (UE). In this context, questions over the transmission of private information over wireless networks naturally arise. To preserve data privacy, a potential solution is to store data on local servers and move network computation to the edge [4], [5]. In fact, data privacy has drawn significant interest in developing new machine learning techniques that can ensure data privacy and exploit the computational resources of users at the same time. One such a promising technique is known as *Federated Learning* (FL) which was first introduced in [6]. In a FL process, base station and users do not share the raw data but only the training updates, and hence, the user privacy is preserved. On the other hand, edge computing does not aim to protect data privacy. Specifically, in edge computing, computations are shifted to the edge devices that are placed close to user devices to reduce the computing burden on user devices. Therefore, there is a risk of privacy leakage as already happened in the third-party companies like Google, Facebook in the past [7]. Moreover, modern user devices are now equipped with more powerful computational capabilities with dedicated and integrated processors like Hexagon DSP with Qualcomm Hexagon Vector eXtensions on Snapdragon 835 [8]. Therefore, computing local updates at user devices in FL is entirely possible. Due to the data privacy attribute, FL has been used in a wide range of real-world digital applications e.g., Gboard, FedVision, functional MRI, FedHealth, etc. [9], [10], [11].

FL has also gained growing attention from the wireless communications research community recently due to its privacy protection and resource utilization features [4], [12], [13], [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], mainly from the viewpoint of implementing FL over wireless networks. The deployment of an FL framework has been studied for 5G and beyond networks [12], [13]. Moreover, pioneering studies on FL in wireless communications can be classified as "learning-oriented" or "communication-oriented." The learning-oriented category aims to improve the learning performance (e.g., training loss, test accuracy) subject to inherent factors in wireless networks such as thermal noise, fading, and estimation errors [15], [16], [17], [18], [19]. Specifically, in [15], Chen et al. considered user selection to minimize the FL training loss function under the presence of network constraints. Amiri and Gündüz in [16] optimized the test accuracy to schedule devices and allocate power across time slots. Sun et al. improved the training efficiency by jointly considering the effects of uplink resource, energy consumption and latency constraints [17]. Bouzinis et al. in [18] considered the problem of minimizing the total delay in each round

Manuscript received 5 April 2022; revised 7 September 2022 and 19 January 2023; accepted 3 May 2023. Date of publication 23 May 2023; date of current version 9 January 2024. This work was supported in part by the Science Foundation Ireland under Grant 17/CDA/4786. The work of Tung Thanh Vu and Hien Quoc Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/X010635/1. The work of Tung Thanh Vu was supported in part by the Excellence Center at Linköping – Lund in Information Technology (ELLIIT) and in part by the Knut and Alice Wallenberg (KAW) Foundation. An earlier version of this paper was presented at the IEEE SPAWC 2022 [DOI: 10.1109/SPAWC51304.2022.9833955]. The associate editor coordinating the review of this article and approving it for publication was Y. Cui. *(Corresponding author: Muhammad Farooq.)*

Muhammad Farooq is with the School of Computer Science, University College Dublin, Dublin 04, D04 V1W8 Ireland (e-mail: muhammad.farooq@ucd.ie).

Tung Thanh Vu is with the Institute of Electronics, Communications and Information Technology, Queen's University Belfast, BT3 9DT Belfast, U.K., and also with the Department of Electrical Engineering (ISY), Linköping University, 581 83 Linköping, Sweden (e-mail: thanh.tung.vu@liu.se).

Hien Quoc Ngo is with the Institute of Electronics, Communications and Information Technology, Queen's University Belfast, BT3 9DT Belfast, U.K. (e-mail: hien.ngo@qub.ac.uk).

Le-Nam Tran is with the School of Electrical and Electronic Engineering, University College Dublin, Dublin 04, D04 V1W8 Ireland (e-mail: nam.tran@ucd.ie).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TWC.2023.3277037.

Digital Object Identifier 10.1109/TWC.2023.3277037

of FL process in the case of compute-then-transmit non-orthogonal multiple access (NOMA). In [19], Wu et al. jointly optimized the overall latency for the FL process and the sum energy consumption of the BS and user devices for the NOMA assisted FL. The communication-oriented category, on the other hand, focuses on enhancing the communication performance (e.g., training latency, energy efficiency) in the framework of FL [20], [21], [22]. For example, in [20], Yang et al. considered the problem of minimization of the total energy consumption to train the FL model under a latency constraint. Vu et al. in [21] focused on minimizing the training latency under transmit power and data rate constraints. In [22], Tran et al. investigated the problem of optimizing the computation and communication latencies of mobile devices subject to various trade-offs between the energy consumption, learning time, and learning accuracy parameters. Zeng et al. in [23] investigated the problem of maximizing the energy efficiency of the FL process by designing a joint computation-and-communication resource management scheme. Pham et al. also studied the energy-efficient resource allocation problem for the FL framework under various resource constraints [24]. All above-mentioned works take into account serving only FL UEs. However, it is certain that future wireless networks will need to serve both the FL and non-FL UEs if FL is to be realized, which calls for novel communication designs. We address this fundamental problem in this paper.

As it is foreseen that future wireless networks will include simultaneously both learning and non-learning users, it will be a natural requirement to serve both group of users at the same time. In this regard, our motivation is to provide an answer to this forward-looking use case. The main challenge of the above coexistence is that how to optimize the data rate of non-FL UEs while ensuring a QoS constraint on the execution time of the FL UEs, which has not yet been studied in the existing literature. To understand this design challenge, let us briefly describe a communication round of an FL iteration in the presence of only FL UEs, which consists of four steps: (i) A central server transmits the global update of an ML model to FL UEs; (ii) FL UEs calculate their local model updates based on their local data set; (iii) The local model updates are sent back to the central server; and (iv) The central server calculates the global update by aggregating the received local model updates [25]. It is clear that problems arise when there are non-FL UEs that need to be served in the downlink. First and most importantly, in Step (iii), the base station needs to set up a two-way communication channel to implement the uplink of FL UEs and the downlink of non-FL UEs. Second, efficient resource allocation approaches are required at all the steps to control the inter-user interference among FL UEs and non-FL UEs to satisfy their different service requirements.

There are two types of communication schemes that are possible to serve the two-way communication between the central server and UEs, namely half-duplex (HD) and full-duplex (FD). Each of these communication schemes has its own advantages and disadvantages [26]. The main draw back of the FD scheme is the self-interference (SI) between transmit and receive antennas of the BS can cause significant performance degradation, which does not appear in the HD communication. However, for small or moderate SI, the FD communication can approximately double the spectral efficiency compared to the half duplex (HD) scheme [27]. Both HD and FD schemes are popular in the literature of massive multiple-input multiple-output (MIMO) networks [28], [29], [30]. However, they cannot be straightforwardly applied to the massive MIMO systems that serve both FL and non-FL UEs.

In this paper, we follow the communication-oriented approach and propose a novel network design for jointly serving FL and downlink non-FL UEs[1] at the same time. First, we propose a communication scheme using massive MIMO and let each FL communication round be executed in one large-scale coherence time.[2] Because of the high array gain, multiplexing gain, and macro-diversity gain, massive MIMO provides a reliable operation of each FL communication round as well as the whole FL process [21]. Here, in the first step of each FL communication round, both groups are jointly served in the downlink by the central server and in the third step, either the HD scheme or the FD scheme is considered to serve the uplink transmission of FL UEs and the downlink transmission of non-FL UEs. Next, we formulate an optimization problem that optimally allocates power and computation resources to maximize the fairness of effective data rates for non-FL UEs, while ensuring a quality-of-service time of each FL iteration for FL UEs. A successive convex approximation algorithm is then proposed to solve the formulated problem. In particular, our contributions are as follows:

- We propose HD and FD communication schemes to jointly serve both FL and non-FL UEs in a massive MIMO network, which has not been studied previously. In the proposed HD scheme, the total system bandwidth is divided equally between the FL and non-FL groups in the uplink of each FL iteration such that both groups are served at the same time in different bandwidths. In the FD communication scheme, both FL UEs and non-FL UEs transmit and receive data in the same time and bandwidth resource under the presence of SI.
- We propose a new performance measure, called the "*effective* data rate", which is defined as the amount of data received by the non-FL UEs, per unit latency time taken by FL UEs. Then, we formulate two optimization problems, each of which is dedicated to the HD and FD scheme, to maximize the minimum effective data subject to a QoS constraint on the execution time for FL UEs. The formulated problems are non-convex with a fractional structure of the objective function and non-convex constraints. To solve the formulated problems, we first transform them into a more tractable form where the SCA is more amendable. Then we apply several convex approximations to arrive at iterative algorithms that are numerically shown to converge very fast. Note that the underlying mathematical structure of the formulated problem is unique and different from any existing works that aim to support federated learning using wireless communications. Therefore, the optimization methods in those works cannot be straightforwardly applied to solve the considered problems in this paper.

[1]The network design for uplink non-FL UEs is open for future works.

[2]Large-scale coherence time is a time interval where the large-scale fading coefficient remains reasonably invariant.

- We provide an extensive set of simulation results to compare the proposed HD-based and FD-based schemes with two baseline schemes: The first baseline scheme makes use of the frequency division multiple access (FDMA) approach to serve each user independently in an allocated bandwidth, while the second baseline scheme considers an equal power allocation (EPA) approach to find the power control. It is observed that the proposed HD and FD schemes provide significantly better solution than two considered baseline schemes. Numerical results also show that the FD scheme is a better choice than the HD scheme when the size of the model updates is large and/or when the SI is small or moderate.

Notations: Bold lower and upper case letters represent vectors and matrices, respectively. The notations $\mathbb{R}$ and $\mathbb{C}$ represent the space of real and complex numbers, respectively. $\|\cdot\|$ represents the Euclidean norm; $|\cdot|$ is the absolute value of the argument. $\mathcal{CN}(0,a)$ denotes a complex Gaussian random variable with zero mean and variance a. $\mathbf{X}^T$ and $\mathbf{X}^H$ stand for the transpose and Hermitian of $\mathbf{X}$, respectively. The operators $\mathbb{E}\{\cdot\}$ and $\mathbb{V}\mathrm{ar}\{\cdot\}$ represent expectation and variance of the argument, respectively.

II. System Model and Proposed Transmission Schemes

A. System Model

We consider a massive MIMO system where a BS serves simultaneously non-FL UEs and FL UEs. We assume that the non-FL UEs are only those receiving data in the downlink transmission.[3] Let $\mathcal{L} \triangleq \{1,\ldots,L\}$, and $\mathcal{K} \triangleq \{1,\ldots,K\}$ be the sets of FL UEs and non-FL UEs, respectively. All FL and non-FL UEs are equipped with a single antenna, while the BS has M transmit antennas and M receive antennas.

To serve FL UEs, the BS acts as a central server. There are four main steps in each iteration of a standard FL framework, i.e., global update downlink transmission, local update computation at the UEs, local update uplink transmission, and global update computation at the BS [6], [22], [31]. To serve non-FL UEs, as mentioned above, the BS constantly transmits downlink data to the non-FL UEs at the same time when all four steps of each FL iteration are executed. Thus, the transmission protocol of our considered system can be summarized as the following four steps in each FL iteration:

(S1) The BS sends a global update through the downlink channel to FL UEs. At the same time, non-FL UEs also receive downlink data from the BS.

(S2) The FL UEs update their local training model based on the global update and solve their local learning problems to obtain their local updates. During this time duration, non-FL UEs continue receiving downlink data from BS.

(S3) The locally computed updates are sent by FL UEs to the BS in the uplink channel while the downlink data is still being sent from the BS to non-FL UEs.

[3]The system model is more general if both uplink and downlink non-FL UEs are included. However, as the first attempt to study a new scenario, we only consider the downlink non-FL UEs to simplify the system model and the resulting mathematical presentation which enables us to obtain important insights. Note that, even for this simplified model, the problem of maximizing the minimum effective rate of non-FL UEs is still non-convex and thus challenging to solve as shall be discussed in the next section. While it is always interesting to consider both uplink and downlink non-FL UEs, we leave it for future research due to the exploring nature of this paper.

TABLE I
TABLE OF FREQUENCY USED NOTATIONS

Notation	Stands For
β_ℓ	Large-scale fading for FL UE ℓ
β_k	Large-scale fading for non-FL UE k
$\mathbf{G}_d$	Channel from BS to FL group in downlink
$\mathbf{H}_i$	Channel from BS to non-FL group in (Si)
$\mathbf{G}_u$	Channel from FL group to BS in uplink
t_d	Downlink transmission time for FL group
t_C	Computation time for FL group in (S2)
t_u	Uplink transmission time for FL group
$\boldsymbol{\eta}_d$	Power coefficients for FL group in downlink
$\boldsymbol{\zeta}_i$	Power coefficients for non-FL group in (Si)
$\boldsymbol{\eta}_u$	Power coefficients for FL group in uplink
$R_{d,\ell}$	Rate for FL UE ℓ in downlink
$R_{i,k}$	Rate for non-FL UE k in (Si)
$R_{u,\ell}$	Rate for FL UE ℓ in uplink

(S4) The BS computes the global update by aggregating the received local updates.

In Step (S3), we need to serve both FL and non-FL UEs. In this regard, there are two types of possible communication schemes: HD and FD. In the HD scheme, the FL and non-FL groups are served in different frequency bands, while in the FD scheme, both groups are served in the same time and frequency resource. During Step (S4), the BS computes its global update after receiving all the local update, the delay of computing the global update is negligible since the computational capability of the central server is much more powerful than that of the UEs. Therefore, the amount of downlink data received by the non-FL UEs during the fourth step is not considered in the rest of the paper. Before proceeding further, we define some frequently used notations in Table I.

B. Proposed Transmission Schemes

In this section, we propose transmission strategies to serve both FL and non-FL UEs at the same time in a massive MIMO network. Such scenario of jointly serving both learning and non-learning group of users has not been addressed in the literature previously. In particular, we propose to use a scheme in [21] to support FL iterations as in Fig. 1(a). We assume that each FL iteration is executed within a large-scale coherence time. All the FL UEs start each step of their FL iterations at the same time, and wait for others to finish their steps before starting a new step. We remark that the small-scale fading coefficients remain constant over each small-scale coherence block, and change in the following coherence blocks. The global and local updates in Steps (S1) and (S3) are transmitted in multiple (small-scale) coherence blocks, as shown in Fig. 1(b). Each small-scale coherence interval in Step (S1) or (S3) includes two phases: channel estimation and downlink or uplink transmission. In the following, we will provide details of our proposed transmission protocol for both HD and HD modes at the BS in Step (S3).

1) Step (S1): In this step, the BS wants to send the global updates to all FL UEs via a downlink transmission while simultaneously sending the payload data to all the non-FL UEs.

- **Channel estimation**: The BS estimates the channels by using uplink pilots received from all the UEs with a time-division-duplexing (TDD) protocol and exploiting channel reciprocity. Let $\sqrt{\tau_{d,p}\rho_p}\boldsymbol{\varphi}_\ell \in \mathbb{C}^{\tau_{d,p}\times 1}$, where

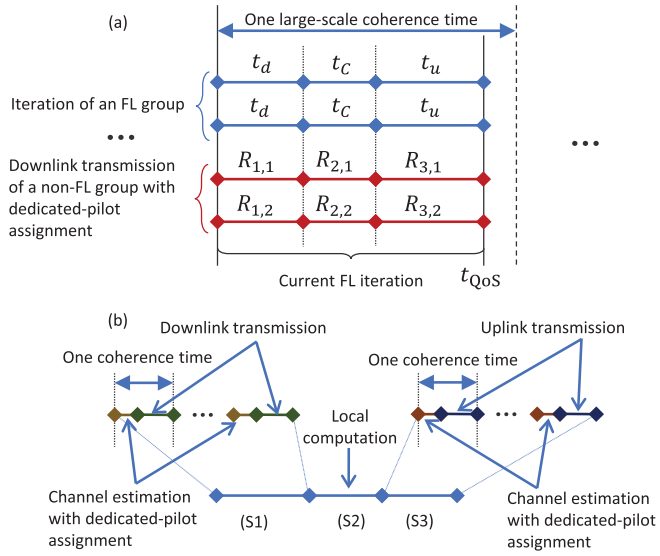


Fig. 1. (a): Illustration of FL iterations over the considered massive MIMO network with two groups of FL and non-FL UEs, and two UEs in each group. (b): Detailed operation of one FL iteration of the FL group.

$\|\boldsymbol{\varphi}_\ell\| = 1$, be the dedicated pilot symbols assigned to the ℓ-th FL UE, and $\sqrt{\tau_{1,p}\rho_p}\bar{\boldsymbol{\varphi}}_k \in \mathbb{C}^{\tau_{1,p}\times 1}$, where $\|\bar{\boldsymbol{\varphi}}_k\| = 1$, be the pilot sequence assigned to the k-th non-FL UE, where ρ_p is the normalized signal to noise ratio (SNR) of each pilot symbol.[4] In addition, $\tau_{d,p}$ and $\tau_{1,p}$ are the corresponding pilot lengths. We assume $\tau_{d,p}, \tau_{1,p} \geq L+K$, and the pilots of non-FL UEs and FL UEs are pairwisely orthogonal i.e. $\boldsymbol{\varphi}_\ell^H \bar{\boldsymbol{\varphi}}_k = 0, \forall \ell, \forall k$, $\boldsymbol{\varphi}_\ell^H \boldsymbol{\varphi}_{\ell'} = 0, \forall \ell' \neq \ell$ and $\bar{\boldsymbol{\varphi}}_k^H \bar{\boldsymbol{\varphi}}_{k'} = 0, \forall k' \neq k$. In other words, pilot contamination is not considered in our paper. Let $\mathbf{G}_d = [\mathbf{g}_{d,1}, \ldots, \mathbf{g}_{d,L}] \in \mathbb{C}^{M\times L}$ and $\mathbf{H}_1 = [\mathbf{h}_{1,1}, \ldots, \mathbf{h}_{1,K}] \in \mathbb{C}^{M\times K}$ be the channel matrices from the BS to the FL and non-FL groups in Step (S1), respectively. Here, $\mathbf{g}_{d,\ell}$ represents the channel vector from the BS to the ℓ-th FL UE, while $\mathbf{h}_{1,k}$ is the channel vector between the BS and non-FL UE k in Step (S1). We assume Rayleigh fading, i.e., $\mathbf{g}_{d,\ell} \sim \mathcal{CN}(\mathbf{0}, \beta_\ell \mathbf{I}_M)$ and $\mathbf{h}_{1,k} \sim \mathcal{CN}(\mathbf{0}, \bar{\beta}_k \mathbf{I}_M)$, where β_ℓ and $\bar{\beta}_k$ represent large-scale fading. The minimum mean square error (MMSE) estimate of $\mathbf{g}_{d,\ell}$ can be written as $\check{\mathbf{g}}_{d,\ell} = \sigma_{d,\ell}\, \mathbf{z}_{d,\ell}$, where $\mathbf{z}_{d,\ell} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, and $\sigma_{d,\ell}^2 = \frac{\rho_p \tau_{d,p} \beta_\ell^2}{\rho_p \tau_{d,p} \beta_\ell + 1}$. Similarly, the MMSE estimate of $\mathbf{h}_{1,k}$ can be written as $\check{\mathbf{h}}_{1,k} = \sigma_{1,k}\, \mathbf{z}_{1,k}$, where $\mathbf{z}_{1,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, and $\sigma_{1,k}^2 = \frac{\rho_p \tau_{1,p} \bar{\beta}_k^2}{\rho_p \tau_{1,p} \bar{\beta}_k + 1}$. Let $\check{\mathbf{G}}_d = [\check{\mathbf{g}}_{d,1}, \ldots, \check{\mathbf{g}}_{d,L}]$, $\check{\mathbf{H}}_1 = [\check{\mathbf{h}}_{1,1}, \ldots, \mathbf{h}_{1,K}]$, $\mathbf{Z}_d = [\mathbf{z}_{d,1}, \ldots, \mathbf{z}_{d,L}]$, $\mathbf{Z}_1 = [\mathbf{z}_{1,1}, \ldots, \mathbf{z}_{1,K}]$, $\boldsymbol{\sigma}_d \triangleq [\sigma_{d,1}, \ldots, \sigma_{d,L}]^T$, and $\boldsymbol{\sigma}_1 \triangleq [\sigma_{1,1}, \ldots, \sigma_{1,K}]^T$. Denote by $\mathbf{E}_d = [\boldsymbol{\epsilon}_{d,1}, \ldots, \boldsymbol{\epsilon}_{d,L}]$ and $\mathbf{E}_1 = [\boldsymbol{\epsilon}_{1,1}, \ldots, \boldsymbol{\epsilon}_{1,K}]$ be the channel estimate error matrices of $\mathbf{G}_d$ and $\mathbf{H}_1$, i.e., $\boldsymbol{\epsilon}_{d,\ell} = \check{\mathbf{g}}_{d,\ell} - \mathbf{g}_{d,\ell}$ and $\boldsymbol{\epsilon}_{1,k} = \check{\mathbf{h}}_{1,k} - \mathbf{h}_{1,k}$. From the property of MMSE estimation, we have that $\boldsymbol{\epsilon}_{d,\ell}$, $\check{\mathbf{g}}_{d,\ell}$, $\boldsymbol{\epsilon}_{1,k}$, and $\check{\mathbf{h}}_{1,k}$ are independent, and hence, $\boldsymbol{\epsilon}_{d,\ell} \sim \mathcal{CN}(\mathbf{0}, (\beta_\ell - \sigma_{d,\ell}^2)\mathbf{I}_M)$, $\boldsymbol{\epsilon}_{1,k} \sim \mathcal{CN}(\mathbf{0}, (\bar{\beta}_k - \sigma_{1,k}^2)\mathbf{I}_M)$.

[4]We follow the practice of normalizing the transmit power to the noise power as done in the literature such as [32, Sec. 3.1.3].

- **Downlink transmission for both FL and non-FL UEs**: The BS encodes downlink data desired for non-FL UE k into the symbol $s_{1,k} \sim \mathcal{CN}(0,1), \forall k \in \mathcal{K}$, and the global training update intended for the FL UE ℓ into symbol $s_{d,\ell} \sim \mathcal{CN}(0,1), \forall \ell \in \mathcal{L}$. Note that the global update is the same for all FL UEs but we use different coding schemes for different UEs. The zero-forcing (ZF) precoding scheme is then applied to precode the symbols for FL and non-FL groups. Let $\mathbf{s}_d \triangleq [s_{d,1}, \ldots, s_{d,L}]^T$, $\mathbf{s}_1 \triangleq [s_{1,1}, \ldots, s_{1,K}]^T$. With ZF, $M \geq L+K$ is required, and the signal transmitted at the BS in Step (S1) is given by

$$\mathbf{x}_1 = \sqrt{\rho_d}\, \mathbf{U}_d\, \mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\, \mathbf{s}_d + \sqrt{\rho_d}\, \mathbf{U}_1\, \mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\, \mathbf{s}_1,$$

where $[\mathbf{U}_d \;\; \mathbf{U}_1] = \sqrt{(M-L-K)}\, \mathbf{Z}(\mathbf{Z}^H \mathbf{Z})^{-1}$ [32, (3.49)], with $\mathbf{Z} = [\mathbf{Z}_d, \mathbf{Z}_1]$. In addition, $\mathbf{D}_{\boldsymbol{\eta}_d}$ and $\mathbf{D}_{\boldsymbol{\zeta}_1}$ are diagonal matrices with the elements of $\boldsymbol{\eta}_d$ and $\boldsymbol{\zeta}_1$ on their diagonal, respectively, where the ℓ-th element of $\boldsymbol{\eta}_d$ denoted by $\eta_{d,\ell}$ and the k-th element of $\boldsymbol{\zeta}_1$ denoted by $\zeta_{1,k}$ are the power control coefficients associated with the ℓ-th FL UE and k-th non-FL UE, respectively. The transmitted power at the BS is required to meet the average normalized power constraint, i.e., $\mathbb{E}\{|\mathbf{x}_1|^2\} \leq \rho_d$, which can be expressed as:

$$\sum\nolimits_{\ell\in\mathcal{L}} \eta_{d,\ell} + \sum\nolimits_{k\in\mathcal{K}} \zeta_{1,k} \leq 1. \tag{1}$$

The received signal vector collected from all FL UEs is given by

$$\begin{aligned}\mathbf{y}_d &= \mathbf{G}_d^H\, \mathbf{x}_1 + \mathbf{n}_d \\ &= \sqrt{\rho_d}\mathbf{G}_d^H\, \mathbf{U}_d\, \mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\, \mathbf{s}_d + \sqrt{\rho_d}\, \mathbf{G}_d^H\, \mathbf{U}_1\, \mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\, \mathbf{s}_1 + \mathbf{n}_d,\end{aligned} \tag{2}$$

where $\mathbf{n}_d \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_L)$ is the additive noise. Since $\check{\mathbf{G}}_d^H\, \mathbf{U}_d = \sqrt{M-L-K}\, \mathbf{D}_{\boldsymbol{\sigma}_d}$ and $\check{\mathbf{G}}_d^H\, \mathbf{U}_1 = \mathbf{0}$, the ℓ-th FL UE receives

$$\begin{aligned}y_{d,\ell} = &\sqrt{\rho_d \eta_\ell (M-L-K)}\sigma_{d,\ell} s_{d,\ell} + n_{d,\ell} \\ &- \sqrt{\rho_d}\boldsymbol{\epsilon}_{d,\ell}^H\, \mathbf{U}_d\, \mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\, \mathbf{s}_d - \sqrt{\rho_d}\boldsymbol{\epsilon}_{d,\ell}^H\, \mathbf{U}_1\, \mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\, \mathbf{s}_1\,.\end{aligned} \tag{3}$$

Following [32, Sec. 3.3.2], the effective SINR at the ℓ-th FL UE is given by

$$\begin{aligned}&\mathrm{SINR}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \\ &= \frac{\rho_d \eta_{d,\ell}(M-L-K)\sigma_{d,\ell}^2}{1 + \rho_d\, \mathbb{V}\mathrm{ar}\{\boldsymbol{\epsilon}_{d,\ell}^H\, \mathbf{U}_d\, \mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\, \mathbf{s}_d + \boldsymbol{\epsilon}_{d,\ell}^H\, \mathbf{U}_1\, \mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\, \mathbf{s}_1\}}.\end{aligned} \tag{4}$$

Since $\boldsymbol{\epsilon}_{d,\ell}$ is independent of $\mathbf{U}_d\, \mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\, \mathbf{s}_d$ and $\mathbf{U}_1\, \mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\, \mathbf{s}_1$, we get the closed-form expression for $\mathrm{SINR}_{d,\ell}$ as

$$\begin{aligned}&\mathrm{SINR}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \\ &= \frac{\rho_d \eta_{d,\ell}(M-L-K)\sigma_{d,\ell}^2}{1 + \rho_d(\beta_\ell - \sigma_{d,\ell}^2)\sum_{i\in\mathcal{L}} \eta_{d,i} + \rho_d(\beta_\ell - \sigma_{d,\ell}^2)\sum_{k\in\mathcal{K}} \zeta_{1,k}}.\end{aligned} \tag{5}$$

Similarly, the received signal vector combined from all the non-FL UEs in Step (S1) is given by

$$\mathbf{y}_1 = \mathbf{H}_1^H\, \mathbf{x}_1 + \mathbf{n}_1$$

$$= \sqrt{\rho_d}\mathbf{H}_1^H\,\mathbf{U}_1\,\mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\,\mathbf{s}_1 + \sqrt{\rho_d}\,\mathbf{H}_1^H\,\mathbf{U}_d\,\mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\,\mathbf{s}_d + \mathbf{n}_1, \tag{6}$$

where $\mathbf{n}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_K)$ is the additive noise. Using the fact that $\check{\mathbf{H}}_1^H\,\mathbf{U}_1 = \sqrt{M-L-K}\,\mathbf{D}_{\boldsymbol{\sigma}_1}$ and $\check{\mathbf{H}}_1^H\,\mathbf{U}_d = \mathbf{0}$, the effective SINR of the k-th non-FL UE is given by

$$\begin{aligned}
&\text{SINR}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \\
&= \frac{\rho_d \zeta_{1,k}(M-L-K)\sigma_{1,k}^2}{1 + \rho_d\,\mathbb{V}\text{ar}\{\boldsymbol{\epsilon}_{1,k}^H\,\mathbf{U}_1\,\mathbf{D}_{\boldsymbol{\zeta}_1}^{1/2}\,\mathbf{s}_1 + \boldsymbol{\epsilon}_{1,k}^H\,\mathbf{U}_d\,\mathbf{D}_{\boldsymbol{\eta}_d}^{1/2}\,\mathbf{s}_d\}} \\
&= \frac{\rho_d \zeta_{1,k}(M-L-K)\sigma_{1,k}^2}{1+\rho_d(\bar{\beta}_k - \sigma_{1,k}^2)\sum_{i\in\mathcal{K}}\zeta_{1,i} + \rho_d(\bar{\beta}_k - \sigma_{1,k}^2)\sum_{\ell\in\mathcal{L}}\eta_{d,\ell}}.
\end{aligned} \tag{7}$$

Since all the FL UEs start and end a step together, the achievable rate (bps) of each FL UE is the minimum achievable rate of the FL group, i.e.,

$$\begin{aligned}
R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) &= \min_{\ell\in\mathcal{L}} R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \\
&\triangleq \min_{\ell\in\mathcal{L}} \frac{\tau_c - \tau_{d,p}}{\tau_c} B \log_2\left(1 + \text{SINR}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)\right),
\end{aligned} \tag{8}$$

where B is the bandwidth and τ_c is the coherence interval. The achievable rate of non-FL UE k is given by

$$R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) = \frac{\tau_c - \tau_{1,p}}{\tau_c} B \log_2\left(1 + \text{SINR}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)\right). \tag{9}$$

- **Downlink delay of the FL group**: Let S_d (bits) be the data size of the global training update of the FL group. The transmission time that the BS needs to send the global update to all FL UEs is given by

$$t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) = \frac{S_d}{R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)}. \tag{10}$$

- **Amount of downlink data received at the non-FL UEs**: The amount of downlink data received at non-FL UE $k \in \mathcal{K}$ is

$$D_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) = R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1). \tag{11}$$

2) Step (S2): After receiving the global update, each FL UE ℓ computes its local training update on its local dataset, while each non-FL UE k keeps receiving its data from the BS.

- **Local computation**: Each FL UE executes N_c local computing rounds over its data set to compute its local update. Let c_ℓ (cycles/sample) be the number of processing cycles for a UE ℓ to process one data sample [22]. Denote by D_ℓ (samples) and f_ℓ (cycles/s) the size of the local data set and the processing frequency of UE ℓ, respectively. To provide a certain synchronization in this step, we choose $f_\ell = \frac{D_\ell c_\ell f}{D_{\max} c_{\max}}$, where $D_{\max} = \max_{\ell\in\mathcal{L}} D_\ell$, $c_{\max} = \max_{\ell\in\mathcal{L}} c_\ell$, and f is a frequency control coefficient. The computation time at all the FL UEs of the FL group is the same $t_C(f)$, which is given by [21], [22]

$$t_C(f) = t_{C,\ell}(f) = \frac{N_c D_\ell c_\ell}{f_\ell} = \frac{N_c D_{\max} c_{\max}}{f}, \forall \ell \in \mathcal{L}. \tag{12}$$

- **Channel estimation for non-FL UEs channel**: In Step (S2), the channel estimation is performed similarly to Step (S1) for the non-FL UEs. The MMSE estimate of $\mathbf{h}_{2,k}$ (the channel between the BS and non-FL UE k in Step (S2) can be written as $\check{\mathbf{h}}_{2,k} = \sigma_{2,k}\,\mathbf{z}_{2,k}$, where $\mathbf{z}_{2,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ and $\sigma_{2,k}^2 = \frac{\rho_p \tau_{2,p} \bar{\beta}_k^2}{\rho_p \tau_{2,p} \bar{\beta}_k + 1}$, where $\tau_{2,p} \geq K$ is the length of pilot sequence in Step (S2).
- **Amount of downlink data received at the non-FL group**: Similarly to Step (S1), ZF is used at the BS to transmit signals to K non-FL UEs. Let $\boldsymbol{\zeta}_2 \triangleq [\zeta_{2,1}, \ldots, \zeta_{2,K}]^T$ be the power control coefficients for non-FL UEs. The transmitted power at the BS is required to meet the average normalized power constraint, which can be expressed as:

$$\sum\nolimits_{k\in\mathcal{K}} \zeta_{2,k} \leq 1. \tag{13}$$

The achievable downlink rate (bps) of non-FL UE $k, \forall k \in \mathcal{K}$, is given by [32, (3.49)]

$$R_{2,k}(\boldsymbol{\zeta}_2) = \frac{\tau_c - \tau_{2,p}}{\tau_c} B \log_2\left(1 + \text{SINR}_{2,k}(\boldsymbol{\zeta}_2)\right), \tag{14}$$

where $\text{SINR}_{2,k}(\boldsymbol{\zeta}_2)$ is the effective SINR given as

$$\text{SINR}_{2,k}(\boldsymbol{\zeta}_2) = \frac{\rho_d \zeta_{2,k}(M-K)\sigma_{2,k}^2}{1 + \rho_d(\bar{\beta}_k - \sigma_{2,k}^2)\sum_{i\in\mathcal{K}}\zeta_{2,i}}. \tag{15}$$

The above equation is similar to (7) except that there is no interference induced by FL UEs in Step (S2). Thus, the total amount of downlink data received at non-FL UE k is

$$D_{2,k}(\boldsymbol{\zeta}_2, f) = R_{2,k}(\boldsymbol{\zeta}_2) t_C(f). \tag{16}$$

3) Step (S3) Using HD:

- **Channel estimation**: In Step (S3), the channels between the BS and FL UEs are estimated using the MMSE estimation technique similarly to Steps (S1) and (S2). The channel $\mathbf{g}_{u,\ell}$ between the BS and FL UEs ℓ in Step (S3) has an estimate $\check{\mathbf{g}}_{u,\ell} = \sigma_{u,\ell}\,\mathbf{z}_{u,\ell}$, where $\mathbf{z}_{u,\ell} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, $\sigma_{u,\ell}^2 = \frac{\rho_p \tau_{u,p} \beta_\ell^2}{\rho_p \tau_{u,p} \beta_\ell + 1}$, and $\tau_{u,p} \geq L + K$ is the pilot length. The channel $\mathbf{h}_{3,k}$ between the BS and the k-th non-FL UEs in Step (S3) has an estimate $\check{\mathbf{h}}_{3,k} = \sigma_{3,k}\,\mathbf{z}_{3,k}$, where $\mathbf{z}_{3,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ and $\sigma_{3,k}^2 = \frac{\rho_p \tau_{3,p} \bar{\beta}_k^2}{\rho_p \tau_{3,p} \bar{\beta}_k + 1}$. Here, $\tau_{3,p} \geq L + K$ is the length of pilot sequence in Step (S3). Let $\mathbf{Z}_u \triangleq [\mathbf{z}_{u,1}, \ldots, \mathbf{z}_{u,L}]$, $\mathbf{Z}_3 \triangleq [\mathbf{z}_{3,1}, \ldots, \mathbf{z}_{3,K}]$, $\mathbf{G}_u \triangleq [\mathbf{g}_{u,1}, \ldots, \mathbf{g}_{u,L}]$, $\check{\mathbf{G}}_u = [\check{\mathbf{g}}_{u,1}, \ldots, \check{\mathbf{g}}_{u,1}]$, $\mathbf{H}_3 \triangleq [\mathbf{h}_{3,1}, \ldots, \mathbf{h}_{3,K}]$, and $\check{\mathbf{H}}_3 = [\check{\mathbf{h}}_{3,1}, \ldots, \check{\mathbf{h}}_{3,K}]$. Denote by $\mathbf{E}_u = [\boldsymbol{\epsilon}_{u,1}, \ldots, \boldsymbol{\epsilon}_{u,L}]$ and $\mathbf{E}_3 = [\boldsymbol{\epsilon}_{3,1}, \ldots, \boldsymbol{\epsilon}_{3,K}]$ be the channel estimate error matrices of $\mathbf{G}_u$ and $\mathbf{H}_3$, i.e., $\mathbf{E}_d = \check{\mathbf{G}}_d - \mathbf{G}_d$ and $\mathbf{E}_3 = \check{\mathbf{H}}_3 - \mathbf{H}_3$. Here, $\boldsymbol{\epsilon}_{u,\ell} \sim \mathcal{CN}(\mathbf{0}, (\beta_\ell - \sigma_{u,\ell}^2)\mathbf{I}_M)$, $\check{\mathbf{g}}_{u,\ell}$, $\boldsymbol{\epsilon}_{3,k} \sim \mathcal{CN}(\mathbf{0}, (\bar{\beta}_k - \sigma_{3,k}^2)\mathbf{I}_M)$, and $\check{\mathbf{h}}_{3,k}$ are independent.
- **Uplink transmission of FL UEs**: After computing the local update, all FL UEs transmit their local updates to the BS. The signal transmitted from FL UE ℓ is

$$x_{u,\ell} = \sqrt{\rho_u \eta_{u,\ell}} s_{u,\ell},$$

where $s_{u,\ell} \sim \mathcal{CN}(0,1)$ is the data symbol, $\eta_{u,\ell}$ is the power control coefficient chosen to satisfy the average

transmit power constraint, i.e., $\mathbb{E}\left\{|x_{u,\ell}|^2\right\} \leq \rho_u$, which can be expressed as

$$\eta_{u,\ell} \leq 1, \quad \forall \ell \in \mathcal{L}\,. \tag{17}$$

The received signal vector at the BS is then given as

$$\mathbf{y}_u^{\text{HD}} = \sqrt{\rho_u}\,\mathbf{G}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \mathbf{n}_u, \tag{18}$$

where $\boldsymbol{\eta}_u = [\eta_{u,1}, \ldots, \eta_{u,L}]^T$ and $\mathbf{n}_u \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the additive noise vector.

After receiving signals from all the UEs, the BS applies a ZF decoding scheme for detecting the FL UEs' symbols. With ZF, signal used for detecting $s_{u,\ell}$ is given by

$$\begin{aligned} y_{u,\ell}^{\text{HD}} &= \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{G}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u \\ &= \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\check{\mathbf{G}}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u - \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u \\ &\quad + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u \\ &= \sqrt{\rho_u \eta_{u,\ell}(M-L)}\,\sigma_{u,\ell}\,s_{u,\ell} - \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u \\ &\quad + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u, \end{aligned} \tag{19}$$

where $\mathbf{u}_{u,\ell} = \sqrt{(M-L)}\,\mathbf{Z}_u(\mathbf{Z}_u^H\,\mathbf{Z}_u)^{-1}\mathbf{e}_{\ell,L}$ is the zero-forcing decoding vector. For synchronization, we choose the rates of FL UEs to be the same as the minimum achievable rates in the FL group, i.e.,

$$\begin{aligned} R_u^{\text{HD}}(\boldsymbol{\eta}_u) &= \min_{\ell \in \mathcal{L}} R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u) \\ &\triangleq \min_{\ell \in \mathcal{L}} \frac{\tau_c - \tau_{u,p}}{\tau_c}\frac{B}{2}\log_2\left(1 + \text{SINR}_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u)\right), \end{aligned} \tag{20}$$

where $1/2$ appears in the pre-log factor of the rate comes from the fact that the system bandwidth is equally divided between the FL and non-FL groups, and

$$\text{SINR}_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u) = \frac{\rho_u \eta_{u,\ell}(M-L)\sigma_{u,\ell}^2}{1 + \rho_d\,\mathbb{V}\text{ar}\{\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u\}}. \tag{21}$$

The above equation is then reduced to

$$\text{SINR}_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u) = \frac{\rho_u \eta_{u,\ell}(M-L)\sigma_{u,\ell}^2}{1 + \rho_u \sum_{i \in \mathcal{L}}(\beta_i - \sigma_{u,i}^2)\eta_{u,i}}. \tag{22}$$

- **Downlink transmission for Non-FL UEs**: Denote by $\mathbf{s}_3 = [s_{3,1} \ \ldots \ s_{3,K}]^T$ the vector of K symbols intended for K non-FL UEs, and $\mathbf{U}_3 = \sqrt{(M-K)}\,\mathbf{Z}_3(\mathbf{Z}_3^H\,\mathbf{Z}_3)^{-1}$ the ZF precoding matrix. Then, the transmitted signal from the BS to the non-FL UEs is given by

$$\mathbf{x}_3 = \sqrt{\rho_d}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3,$$

where $\boldsymbol{\zeta}_3 \triangleq [\zeta_{3,1}, \ldots, \zeta_{3,K}]^T$, and $\zeta_{3,k}$ the power control coefficient allocated for non-FL UE k chosen to meet the average normalized power constraint at the BS, i.e., $\mathbb{E}\{|\,\mathbf{x}_3\,|^2\} \leq \rho_d$, which can be expressed as:

$$\sum_{k \in \mathcal{K}} \zeta_{3,k} \leq 1. \tag{23}$$

For the k-th non-FL UE, the received signal can be written as

$$\begin{aligned} y_{3,k}^{\text{HD}} &= \sqrt{\rho_d}\check{\mathbf{h}}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 - \sqrt{\rho_d}\,\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + n_{3,k} \\ &= \sqrt{\rho_d \eta_{3,k}(M-K)}\,\sigma_{3,k}\,s_{3,k} \\ &\quad - \sqrt{\rho_d}\,\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + n_{3,k}. \end{aligned} \tag{24}$$

In the above equation, the term $\boldsymbol{\epsilon}_{3,k}$ is independent of $\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_u}^{1/2}\,\mathbf{s}_3$. Thus, under the HD scheme in Step (S3), the effective SINR for the downlink payload at non-FL UE k is

$$\begin{aligned} \text{SINR}_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3) &= \frac{\rho_d \zeta_{3,k}(M-K)\sigma_{3,k}^2}{1 + \rho_d\,\mathbb{V}\text{ar}\{\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3\}} \\ &= \frac{\rho_d \zeta_{3,k}(M-K)\sigma_{3,k}^2}{1 + \rho_d(\bar{\beta}_k - \sigma_{3,k}^2)\sum_{i \in \mathcal{K}}\zeta_{3,i}}, \end{aligned} \tag{25}$$

and the achievable downlink rate for non-FL UE $k, \forall k \in \mathcal{K}$, is

$$R_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3) = \frac{\tau_c - \tau_{3,p}}{\tau_c}\frac{B}{2}\log_2\left(1 + \text{SINR}_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3)\right). \tag{26}$$

- **Uplink delay**: Denote by S_u (bits) the data size of the local training update of the FL group. The transmission time from each FL UE to the BS is the same and given by

$$t_u^{\text{HD}}(\boldsymbol{\eta}_u) = \frac{S_u}{R_u^{\text{HD}}(\boldsymbol{\eta}_u)}. \tag{27}$$

- **Amount of downlink data received at the non-FL group**: The amount of downlink data received at the non-FL UE $k, \forall k \in \mathcal{K}$, in Step (S3) using HD is

$$D_{3,k}^{\text{HD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = R_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3)\,t_u^{\text{HD}}(\boldsymbol{\eta}_u). \tag{28}$$

Before proceeding further, we remark that in Step (S3) using HD, the uplink transmission of FL users and the downlink transmission of non-FL users need to be served at the same time. However, a traditional half-duplex base station only operates in either uplink mode or downlink mode, which raises a challenge for the system design. Therefore, a simple approach is to divide the bandwidth equally for the half-duplex base station to serve each user group in each half of the bandwidth, as done in our paper. However, even with the equal bandwidth allocation, as shall be seen shortly, the problem is already very complicated and challenging to solve. We also remark that looking from an algorithmic viewpoint, our proposed solution for the HD case to be presented in the next section is also applicable to any fixed bandwidth allocation. However, finding the optimal bandwidth allocation will make the resulting problem more intractable, which is out of the scope of the paper and thus will be left for future work.

4) Step (S3) Using FD: Step (S3) involves transmission in both directions. This motivates us to consider the FD communications to serve both groups of UEs simultaneously. Specifically, FL UEs send their local updates to the BS in the uplink channel and at the same time, non-FL UEs receive the downlink data from the BS. The proposed FD scheme is detailed in what follows.

- **Uplink transmission of FL UEs**: In the FD communications, channel coefficients are estimated similarly to what was done in the case of the HD communications. FL UEs transmit the locally computed updates to the BS in the presence of non-FL UEs which are receiving the downlink data. Therefore, SI is present between the

receiver and transmit antennas of the BS which is denoted by $\mathbf{G}^{\text{SI}} \in \mathbb{C}^{M\times M}$. The elements of matrix $\mathbf{G}_{\text{SI}}$ are modeled as i.i.d. random variables with a variance being given by $\sigma^2_{\text{SI}} = \beta_{\text{SI}}\sigma^2_{\text{SI},0}$, where β_{SI} represents the path loss from a transmit antenna to a receive antenna of the BS due to their physical antenna separation and $\sigma^2_{\text{SI},0}$ is the power of the residual interference at each BS antenna after the SI suppression, respectively. Similar to the HD scheme, the baseband signal is subjected to the average transmit power constraint (17). The received signal vector at the BS in the FD communication is expressed as

$$\mathbf{y}_u^{\text{FD}} = \sqrt{\rho_u}\,\mathbf{G}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \sqrt{\rho_d}\,\mathbf{G}_{\text{SI}}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \mathbf{n}_u, \tag{29}$$

where $\mathbf{n}_u \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the vector of additive noise components. Note that SI is caused from transmit antennas of the BS to receiving antennas and thus, the effective noise has an additional SI term caused by the downlink transmission to non-FL UEs. After receiving signals from all the UEs, the BS applies a ZF decoding scheme for detecting the FL UEs' symbols. The detected signal for the ℓ-th FL UE is given in (30), as shown at the bottom of the page. The SINR for the ℓ-th FL UE in case of FD communications is given as

$$\text{SINR}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_u \eta_{u,\ell}(M-L)\sigma^2_{u,\ell}}{1+\rho_d\,\mathbb{V}\text{ar}\{\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \mathbf{u}_{u,\ell}^H\,\mathbf{G}_{\text{SI}}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_u\}}. \tag{31}$$

Proposition 1: The SINR for the ℓ-th FL UE in case of FD communications given in (31) can be approximated by

$$\text{SINR}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \approx \widehat{\text{SINR}}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_u \eta_{u,\ell}(M-L)\sigma^2_{u,\ell}}{1+\rho_u\sum_{i\in\mathcal{L}}(\beta_i - \sigma^2_{u,i})\eta_{u,i} + \rho_d M \beta_{\text{SI}}\sigma^2_{\text{SI},0}\sum_{j\in\mathcal{K}}\zeta_{3,j}}. \tag{32}$$

Proof: See Appendix A. ∎

For synchronization, we again choose the rates of FL UEs to be the same as the minimum achievable rates in the FL group.

$$R_u^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \min_{\ell\in\mathcal{L}} R_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \triangleq \min_{\ell\in\mathcal{L}} \frac{\tau_c - \tau_{u,p}}{\tau_c} B \log_2\left(1 + \widehat{\text{SINR}}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right). \tag{33}$$

The above equation is similar to (20) except that FL UEs make use of the full bandwidth in the FD communication.

- **Downlink transmission for Non-FL UEs**: In FD, non-FL UEs continue receiving data from the BS in the downlink channel in the presence of FL UEs which simultaneously send the local updates to the BS in the uplink channel. Therefore, the received signal at each non-FL UE contains the inter-group interference (IGI) from the group of FL UEs. To approximate the SINR in this case, the transmitted power at the BS is constrained to meet the average normalized power constraint (23) similar to the HD scheme.

The received signal for the k-th non-FL UE can be written as

$$y_{3,k}^{\text{FD}} = \sqrt{\rho_d}\,\mathbf{h}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \sqrt{\rho_u}\,\mathbf{H}_{\text{IGI}}\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + n_{3,k}, \tag{34}$$

where $\mathbf{H}_{\text{IGI}} \in \mathbb{C}^{L\times K}$ is the inter-group channel matrix whose elements are modeled as $h_{\text{IGI},k\ell} = \beta_{\text{IGI},k\ell}^{1/2}\bar{h}_{\text{IGI},k\ell}$, where $\beta_{\text{IGI},k\ell}$ is the large-scale fading and $\bar{h}_{\text{IGI},k\ell} \sim \mathcal{CN}(0,1)$ is the small-scale fading of the inter-group channel. After the channel estimation, the first term in the above equation can be broken into the estimation term and the error term and thus, the above equation can be rewritten as given in (35), as shown at the bottom of this page. The effective SINR in the downlink payload at non-FL UE k is given by

$$\text{SINR}_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_d \eta_{3,k}(M-K)\sigma^2_{3,k}}{1+\mathbb{V}\text{ar}\{\rho_d\,\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \rho_u\,\mathbf{H}_{\text{IGI},k}^H\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u\}}. \tag{36}$$

Note that in the above equation, $\boldsymbol{\epsilon}_{3,k}$ is independent of $\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_u}^{1/2}\,\mathbf{s}_3$. Moreover, $\mathbb{V}\text{ar}\{\rho_u\,\mathbf{H}_{\text{IGI},k}^H\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u\}$ simplifies to $\rho_u\sum_{i\in\mathcal{L}}\eta_{u,i}\beta_{\text{IGI},ki}$. Thus, the effective SINR can be rewritten as

$$\text{SINR}_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_d \eta_{3,k}(M-K)\sigma^2_{3,k}}{1+\rho_d(\bar{\beta}_k - \sigma^2_{3,k})\sum_{j\in\mathcal{K}}\zeta_{3,j} + \rho_u\sum_{i\in\mathcal{L}}\eta_{u,i}\beta_{\text{IGI},ki}}. \tag{37}$$

Now, the achievable downlink rate for non-FL UE $k, \forall k \in \mathcal{K}$, is

$$R_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\tau_c - \tau_{3,p}}{\tau_c} B \log_2\left(1 + \text{SINR}_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right). \tag{38}$$

$$\begin{aligned} y_{u,\ell}^{\text{FD}} &= \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{G}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \sqrt{\rho_d}\,\mathbf{u}_{u,\ell}^H\,\mathbf{G}_{\text{SI}}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u \\ &= \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\check{\mathbf{G}}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u - \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \sqrt{\rho_d}\,\mathbf{u}_{u,\ell}^H\,\mathbf{G}_{\text{SI}}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u \\ &= \sqrt{\rho_u\eta_{u,\ell}(M-L)}\sigma_{u,\ell}s_{u,\ell} - \sqrt{\rho_u}\,\mathbf{u}_{u,\ell}^H\,\mathbf{E}_u\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + \sqrt{\rho_d}\,\mathbf{u}_{u,\ell}^H\,\mathbf{G}_{\text{SI}}\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \mathbf{u}_{u,\ell}^H\,\mathbf{n}_u\,. \end{aligned} \tag{30}$$

$$\begin{aligned} y_{3,k}^{\text{FD}} &= \sqrt{\rho_d}\check{\mathbf{h}}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 - \sqrt{\rho_d}\,\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \sqrt{\rho_u}\,\mathbf{H}_{\text{IGI}}\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + n_{3,k} \\ &= \sqrt{\rho_d\eta_{3,k}(M-K)}\sigma_{3,k}s_{3,k} - \sqrt{\rho_d}\,\boldsymbol{\epsilon}_{3,k}^H\,\mathbf{U}_3\,\mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2}\,\mathbf{s}_3 + \sqrt{\rho_u}\,\mathbf{H}_{\text{IGI}}\,\mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\,\mathbf{s}_u + n_{3,k}. \end{aligned} \tag{35}$$

- **Uplink delay**: Denote by S_u (bits) the data size of the local training update of the FL group. The transmission time from FL UE ℓ to the BS is the same and given by

$$t_u^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{S_u}{R_u^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)}. \tag{39}$$

  The above equation is similar to (27) except that the transmission time now depends on power control coefficients from both FL and non-FL UEs.
- **Amount of downlink data received at the non-FL group**: The amount of downlink data received at all non-FL UE $k, \forall k \in \mathcal{K}$, in Step (S3) is

$$D_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = R_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) t_u^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3). \tag{40}$$

  This equation is also similar to (28) while the only difference is that the downlink rate of k-th non-FL UE and the transmission time of the FL UEs depend on both $\boldsymbol{\eta}_u$ and $\boldsymbol{\zeta}_3$.

5) Step (S4): After receiving all the local update, the BS (i.e., central server) computes its global update. Since the computational capability of the central server is much more powerful than that of the UEs, the delay of computing the global update is negligible.

III. Problem Formulation and Proposed Solution

The problem of fairness among the non-FL UEs in terms of effective data received is one of the key challenges in wireless communications. In this section we first define a new performance metric which is referred to as the effective data rate of non-FL UEs and then formulate the optimization problems to achieve the max-min fairness of non-FL UEs subject to a QoS constraint on the execution time of FL UEs.

A. Effective Data Rate of Non-FL UEs

From the discussions in the preceding section, the data rate of each non-FL UE is changed for different steps. Thus, it is practically reasonable to use the average data rate accounting for all steps as a representative data rate for the system design purposes. More specifically, the total amount of data received by the k-th non-FL UE in Steps (S1)-(S3) is $D_{1,k} + D_{2,k} + D_{3,k}^{\text{mode}}$, where mode $\in \{\text{HD}, \text{FD}\}$. Also, the time of each step is determined by the FL UEs. It is obvious that the total time of the three steps is $t_d + t_C + t_u^{\text{mode}}$. Thus, we define the effective data rate for the k-th non-FL UE as $\frac{D_{1,k}+D_{2,k}+D_{3,k}^{\text{mode}}}{t_d+t_C+t_u^{\text{mode}}}$. In the following we use this definition of the effective data rate for non-FL UEs to formulate max-min fairness problems for HD and FD approaches.

B. HD Scheme

1) Problem Formulation for HD Scheme: The considered problem for the HD communication scheme can be mathematically expressed as follows:

$$\max_{\mathbf{x}^{\text{HD}}} \frac{\min_{k\in\mathcal{K}} \left(D_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + D_{2,k}(f, \boldsymbol{\zeta}_2) + D_{3,k}^{\text{HD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right)}{t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t_u^{\text{HD}}(\boldsymbol{\eta}_u)} \tag{41a}$$

$$\text{s.t. } (1), (13), (17), (23),$$

$$\eta_{d,\ell}, \quad \zeta_{1,k}, \ \zeta_{2,k}, \ \eta_{u,\ell}, \ \zeta_{3,k} \geq 0, \ \eta_{d,\ell} \leq 1, \tag{41b}$$

$$f_{\min} \leq f_\ell \leq f_{\max}, \quad \forall \ell \tag{41c}$$

$$t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t_u^{\text{HD}}(\boldsymbol{\eta}_u) \leq t_{\text{QoS}}^{\text{HD}}, \tag{41d}$$

where $\mathbf{x}^{\text{HD}} \triangleq [\boldsymbol{\eta}_d^T, \boldsymbol{\zeta}_1^T, f, \boldsymbol{\zeta}_2^T, \boldsymbol{\eta}_u^T, \boldsymbol{\zeta}_3^T]^T$. The constraint (41d) is introduced to ensure that the time taken by the FL UEs is bounded by $t_{\text{QoS}}^{\text{HD}}$.

2) Solution for HD Scheme: We now present a solution to (41) based on successive convex approximation (SCA). Our idea is to equivalent transform the sophisticated constraints into simpler ones where convex approximations are easier to find. To this end, using the epigraph form, we first equivalently rewrite (41) as

$$\max_{\bar{\mathbf{x}}^{\text{HD}}} \frac{t^{\text{HD}}}{t_{\text{Q}}^{\text{HD}}} \tag{42a}$$

$$\text{s.t. } (1), (13), (17), (23), (41b), (41c),$$

$$t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t_u^{\text{HD}}(\boldsymbol{\eta}_u) \leq t_{\text{Q}}^{\text{HD}}, \tag{42b}$$

$$D_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + D_{2,k}(f, \boldsymbol{\zeta}_2) + D_{3,k}^{\text{HD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \geq t^{\text{HD}}, \quad \forall k \tag{42c}$$

$$t_{\text{Q}}^{\text{HD}} \leq t_{\text{QoS}}^{\text{HD}}, \tag{42d}$$

where $\bar{\mathbf{x}}^{\text{HD}} = [(\mathbf{x}^{\text{HD}})^T, t^{\text{HD}}, t_{\text{Q}}^{\text{HD}}]^T$. Next, it is straightforward to see that (42) can be further equivalently reformulated as

$$\max_{\bar{\mathbf{x}}^{\text{HD}}} \frac{t^{\text{HD}}}{t_{\text{Q}}^{\text{HD}}} \tag{43a}$$

$$\text{s.t. } (1), (13), (17), (23), (41b), (41c), (42d),$$

$$\frac{S_d}{R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)} + \frac{N_c D_{\max} c_{\max}}{f} + \frac{S_u}{R_u^{\text{HD}}(\boldsymbol{\eta}_u)} \leq t_{\text{Q}}^{\text{HD}}, \tag{43b}$$

$$R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \frac{S_d}{R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)} + R_{2,k}(\boldsymbol{\zeta}_2) \frac{N_c D_{\max} c_{\max}}{f} + R_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3) \frac{S_u}{R_u^{\text{HD}}(\boldsymbol{\eta}_u)} \geq t^{\text{HD}}, \quad \forall k. \tag{43c}$$

It is now clear that (43b) and (43c) are troublesome. We note that (43b) is equivalent to the following set of constraints

$$\frac{S_d}{r_d} + \frac{N_c D_{\max} c_{\max}}{f} + \frac{S_u}{r_u^{\text{HD}}} \leq t_{\text{Q}}^{\text{HD}}, \tag{44a}$$

$$r_d \leq R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall \ell \tag{44b}$$

$$r_u^{\text{HD}} \leq R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u), \quad \forall \ell \tag{44c}$$

$$r_d \geq 0, \quad r_u^{\text{HD}} \geq 0. \tag{44d}$$

It is easy to see that (43b) and (44) are equivalent since if $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u)$ are feasible to (43b), then they are also feasible to (44) and vice versa. We note that (44a) is convex. Intuitively, r_d and r_u^{HD} are lower-bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u)$, respectively, for all $\ell \in \mathcal{L}$. In the same way, to deal with (43c), we rewrite it as

$$r_{1,k} \frac{S_d}{\tilde{r}_d} + r_{2,k} \frac{N_c D_{\max} c_{\max}}{f} + r_{3,k}^{\text{HD}} \frac{S_u}{\tilde{r}_u^{\text{HD}}} \geq t^{\text{HD}}, \quad \forall k \tag{45a}$$

$$R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \leq \tilde{r}_d, \quad \forall \ell \tag{45b}$$

$$R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u) \leq \tilde{r}_u^{\text{HD}}, \quad \forall \ell, \tag{45c}$$

$$r_{1,k} \leq R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall k \tag{45d}$$

$$r_{2,k} \leq R_{2,k}(\boldsymbol{\zeta}_2), \quad \forall k \tag{45e}$$

$$r_{3,k}^{\mathrm{HD}} \le R_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3), \quad \forall k \tag{45f}$$

$$\tilde{r}_d \ge 0, \quad \tilde{r}_u \ge 0, \; r_{1,k} \ge 0, \; r_{2,k} \ge 0, \; r_{3,k}^{\mathrm{HD}} \ge 0, \; \forall k \tag{45g}$$

where $\tilde{r}_d$ and $\tilde{r}_u^{\mathrm{HD}}$ are respectively seen as upper-bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $R_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u)$ for all $\ell \in \mathcal{L}$, and $r_{1,k}$, $r_{2,k}$ and $r_{3,k}^{\mathrm{HD}}$ the lower-bounds of $R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $R_{2,k}(\boldsymbol{\zeta}_2)$ and $R_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3)$. We can now further equivalently express (45a) as

$$t^{\mathrm{HD}} \le a_1 S_d + a_2 N_c D_{\max} c_{\max} + a_3^{\mathrm{HD}} S_u, \tag{46a}$$

$$a_1 \le \frac{r_{1,k}}{\tilde{r}_d} \Leftrightarrow a_1 \tilde{r}_d \le r_{1,k}, \quad \forall k \tag{46b}$$

$$a_2 \le \frac{r_{2,k}}{f} \Leftrightarrow a_2 f \le r_{2,k}, \quad \forall k \tag{46c}$$

$$a_3^{\mathrm{HD}} \le \frac{r_{3,k}^{\mathrm{HD}}}{\tilde{r}_u^{\mathrm{HD}}} \Leftrightarrow a_3^{\mathrm{HD}} \tilde{r}_u^{\mathrm{HD}} \le r_{3,k}^{\mathrm{HD}}, \quad \forall k. \tag{46d}$$

The above transformations mean (41) is equivalent to the following problem

$$\max_{\tilde{\mathbf{x}}^{\mathrm{HD}}} \; z^{\mathrm{HD}} \tag{47a}$$

$$\text{s.t. } z^{\mathrm{HD}} t_{\mathrm{Q}}^{\mathrm{HD}} \le t^{\mathrm{HD}}, \; (1), (13), (17), (23), (41b), (41c), (42d), (44), (45b) - (45f), (46), \tag{47b}$$

where $\tilde{\mathbf{x}}^{\mathrm{HD}} \triangleq [(\bar{\mathbf{x}}^{\mathrm{HD}})^T, r_d, r_u^{\mathrm{HD}}, a_1, a_2, a_3^{\mathrm{HD}}, \mathbf{r}_1^T, \mathbf{r}_2^T, (\mathbf{r}_3^{\mathrm{HD}})^T, \tilde{r}_d, \tilde{r}_u^{\mathrm{HD}}, z^{\mathrm{HD}}]^T$, $\mathbf{r}_1 = \{r_{1,k}\}$, $\mathbf{r}_2 = \{r_{2,k}\}$, $\mathbf{r}_3^{\mathrm{HD}} = \{r_{3,k}^{\mathrm{HD}}\}$.

Problem (47) is still difficult to solve due to nonconvex constraints (44b)-(45f), (46b)-(46d), and (47b). However, these constraints are amenable to applying the SCA method, which we show next.

In the sequel, we denote $\tilde{\mathbf{x}}^{\mathrm{HD}(n)}$ to be the value of $\tilde{\mathbf{x}}^{\mathrm{HD}}$ after n iterations. We first note that constraints (44b), (44c), (45d)–(45f) are of the same type in the sense that concave lower bounds of the involving rate expressions are required to obtain convex approximate constraints. To this end we recall the following inequality

$$\log\left(1 + \frac{x}{y}\right) \ge \log\left(1 + \frac{x^{(n)}}{y^{(n)}}\right) + \frac{2x^{(n)}}{(x^{(n)} + y^{(n)})} - \frac{(x^{(n)})^2}{(x^{(n)} + y^{(n)})x} - \frac{x^{(n)} y}{(x^{(n)} + y^{(n)})y^{(n)}}, \tag{48}$$

where $x > 0, y > 0$ [33, (76)]. Applying the above inequality we obtain the following inequalities

$$\tilde{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \le R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall \ell \tag{49a}$$

$$\tilde{R}_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u) \le R_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u), \quad \forall \ell \tag{49b}$$

$$\tilde{R}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \le R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall k \tag{49c}$$

$$\tilde{R}_{2,k}(\boldsymbol{\zeta}_2) \le R_{2,k}(\boldsymbol{\zeta}_2), \quad \forall k \tag{49d}$$

$$\tilde{R}_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3) \le R_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3), \quad \forall k \tag{49e}$$

where $\tilde{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $\tilde{R}_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u)$, $\tilde{R}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $\tilde{R}_{2,k}(\boldsymbol{\zeta}_2)$, and $\tilde{R}_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3)$ are concave lower bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $R_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u)$, $R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $R_{2,k}(\boldsymbol{\zeta}_2)$, and $R_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3)$, respectively. The expressions of these lower bounds are given in (75) in Appendix C. Consequently, in light of SCA, (44b), (44c), (45d)–(45f) are approximated by the following convex constraints

$$r_d \le \tilde{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall \ell \tag{50a}$$

$$r_u^{\mathrm{HD}} \le \tilde{R}_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u), \quad \forall \ell \tag{50b}$$

$$r_{1,k} \le \tilde{R}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall k \tag{50c}$$

$$r_{2,k} \le \tilde{R}_{2,k}(\boldsymbol{\zeta}_2), \quad \forall k \tag{50d}$$

$$r_{3,k}^{\mathrm{HD}} \le \tilde{R}_{3,k}^{\mathrm{HD}}(\boldsymbol{\zeta}_3), \quad \forall k. \tag{50e}$$

To proceed further we note that constraints (46b)–(46d) and (47b) are of the same type. To deal with these, let us recall the following equality

$$xy = \frac{1}{4}[(x+y)^2 - (x-y)^2]. \tag{51}$$

Since we need a *convex upper bound* of the term xy, a simple way is to linearize the term $(x-y)^2$.

$$xy \le \frac{1}{4}[(x+y)^2 - 2(x^{(n)} - y^{(n)})(x-y) + (x^{(n)} - y^{(n)})^2], \tag{52}$$

where $x \ge 0, y \ge 0$, and $x^{(n)}$ and $y^{(n)}$ are the values of x and y at the n-th iteration, respectively [21]. Thus, using (52) we can approximate (46b)–(46d) and (47b) by the following convex constraints, respectively.

$$\frac{1}{4}[(a_1 + \tilde{r}_d)^2 - (a_1^{(n)} - \tilde{r}_d^{(n)})^2 + 2(a_1 - \tilde{r}_d) \times (a_1^{(n)} - \tilde{r}_d^{(n)})] \le r_{1,k}, \quad \forall k \tag{53a}$$

$$\frac{1}{4}[(a_2 + f)^2 - (a_2^{(n)} - f^{(n)})^2 + 2(a_2 - f) \times (a_2^{(n)} - f^{(n)})] \le r_{2,k}, \quad \forall k \tag{53b}$$

$$\frac{1}{4}\big[(a_3^{\mathrm{HD}} + \tilde{r}_u^{\mathrm{HD}})^2 - (a_3^{\mathrm{HD}(n)} - \tilde{r}_u^{\mathrm{HD}(n)})^2 + 2(a_3^{\mathrm{HD}} - \tilde{r}_u^{\mathrm{HD}}) \times (a_3^{\mathrm{HD}(n)} - \tilde{r}_u^{\mathrm{HD}(n)})\big] \le r_{3,k}^{\mathrm{HD}}, \quad \forall k \tag{53c}$$

$$\frac{1}{4}\big[(z^{\mathrm{HD}} + t_{\mathrm{Q}}^{\mathrm{HD}})^2 - (z^{\mathrm{HD}(n)} - t_{\mathrm{Q}}^{\mathrm{HD}(n)})^2 + 2(z^{\mathrm{HD}} - t_{\mathrm{Q}}^{\mathrm{HD}}) \times (z^{\mathrm{HD}(n)} - t_{\mathrm{Q}}^{\mathrm{HD}(n)})\big] \le t^{\mathrm{HD}}. \tag{53d}$$

We now turn our attention to (45b) and (45c). It is obvious now we need to derive *convex upper bounds* of the rate functions present in these two constraints. To this end we resort to the following inequality

$$\log\left(1 + \frac{x}{y}\right) \le \log\left(1 + \frac{x^{(n)}}{y^{(n)}}\right) + \frac{y^{(n)}}{(x^{(n)} + y^{(n)})} \times \left(\frac{(x^2 + (x^{(n)})^2)}{2x^{(n)} y} - \frac{x^{(n)}}{y^{(n)}}\right), \tag{54}$$

where $x > 0, y > 0$, and $x^{(n)}$ and $y^{(n)}$ are the values of x and y at the n-th iteration, respectively [34, (75)]. Using this inequality we can approximate constraints (45b) and (45c) by the following convex constraints

$$\hat{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \le \tilde{r}_d, \quad \forall \ell \tag{55a}$$

$$\hat{R}_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u) \le \tilde{r}_u^{\mathrm{HD}}, \quad \forall \ell, \tag{55b}$$

Algorithm 1 Algorithm for Solving (41)

1: Input: Set $n=0$ and choose an initial point $\tilde{\mathbf{x}}^{\mathrm{HD}(0)} \in \tilde{\mathcal{F}}^{\mathrm{HD}}$
2: **repeat**
3: Solve (56) to get $\tilde{\mathbf{x}}^{\mathrm{HD}*}$
4: $\tilde{\mathbf{x}}^{\mathrm{HD}(n+1)} \leftarrow \tilde{\mathbf{x}}^{\mathrm{HD}*}$
5: $n \leftarrow n+1$
6: **until** convergence

where $\hat{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $\hat{R}^{\mathrm{HD}}_{u,\ell}(\boldsymbol{\eta}_u)$ are convex upper bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $R^{\mathrm{HD}}_{u,\ell}(\boldsymbol{\eta}_u)$, respectively. The detailed expressions of $\hat{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $\hat{R}^{\mathrm{HD}}_{u,\ell}(\boldsymbol{\eta}_u)$ are given in (76) in Appendix C.

In summary, at iteration $n+1$, problem (43) is approximated by the following convex problem:

$$\max \{z^{\mathrm{HD}} \mid \tilde{\mathbf{x}}^{\mathrm{HD}} \in \tilde{\mathcal{F}}^{\mathrm{HD}}\}, \tag{56}$$

where $\tilde{\mathcal{F}}^{\mathrm{HD}} \triangleq \{(1), (13), (17), (23), (41b), (41c), (42d), (44a), (44d), (45g), (46a), (50), (53), (55)\}$. We outline the main steps to solve problem (47) in Algorithm 1.

Remark 1: Algorithm 1 requires a feasible point to start the iterative procedure. In general, it is difficult to find a feasible solution to (47). We now describe a practical way to overcome this issue. It is not difficult to see that by randomly generating and properly the variables in $\mathbf{x}^{\mathrm{HD}}$ we can meet (1), (13), (17), (23), (41b), (41c). The remaining variables in $\tilde{\mathbf{x}}^{\mathrm{HD}}$ can be found by letting the corresponding inequality constraint (47) be binding (i.e., occur with equality). If (42d) is satisfied, then we can use this initial point to start Algorithm 1. When the requirements are high (e.g., when $t^{\mathrm{HD}}_{\mathrm{QoS}}$ is small), it is likely that (42d) is not met. In such cases, we introduce a slack variable s and replacing (56) by the following problem

$$\max_{s\le 0, \tilde{\mathbf{x}}^{\mathrm{HD}}} \; z^{\mathrm{HD}} + \alpha s \tag{57a}$$

$$\text{s.t. } \tilde{\mathbf{x}}^{\mathrm{HD}} \in \tilde{\mathcal{F}}^{\mathrm{HD}} \setminus (42d), \tag{57b}$$

$$t^{\mathrm{HD}}_{\mathrm{Q}} + s \le t^{\mathrm{HD}}_{\mathrm{QoS}}. \tag{57c}$$

Intuitively, s represents the violation of (42d) and $\alpha > 0$ is the penalty parameter. It is easy to see that (57c) is met if s is sufficiently small, and thus (57) is always feasible. On the other hand, the maximization of the regularized objective in (57a) will force s to approach 0 when the iterative process progresses. Thus, when Algorithm 1 converges and if $|s|$ is smaller than a pre-determined error tolerance, we will take $\tilde{\mathbf{x}}^{\mathrm{HD}*}$ as the final solution. Otherwise, we say that (41) is infeasible.

C. FD Scheme

1) Problem Formulation for FD Communication Scheme: The considered problem for the FD communication scheme is mathematically stated as

$$\max_{\mathbf{x}^{\mathrm{FD}}} \frac{\min_{k\in\mathcal{K}} \left(D_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + D_{2,k}(f, \boldsymbol{\zeta}_2) + D^{\mathrm{FD}}_{3,k}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right)}{t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t^{\mathrm{FD}}_u(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)} \tag{58a}$$

$$\text{s.t. } (1), (13), (17), (23),$$

$$\eta_{d,\ell}, \quad \zeta_{1,k}, \; \zeta_{2,k}, \; \eta_{u,\ell}, \; \zeta_{3,k} \ge 0, \; \eta_{d,\ell} \le 1, \tag{58b}$$

$$f_{\min} \le f_\ell \le f_{\max}, \quad \forall \ell \tag{58c}$$

$$t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t^{\mathrm{FD}}_u(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \le t^{\mathrm{FD}}_{\mathrm{QoS}}, \tag{58d}$$

where $\mathbf{x}^{\mathrm{FD}} \triangleq [\boldsymbol{\eta}_d^T, \boldsymbol{\zeta}_1^T, f, \boldsymbol{\zeta}_2^T, \boldsymbol{\eta}_u^T, \boldsymbol{\zeta}_3^T]^T$.

2) Solution for FD Scheme: The solution for the FD scheme follows closely the derivations of that for the HD scheme. First, we equivalently rewrite (58) as

$$\max_{\bar{\mathbf{x}}^{\mathrm{FD}}} \frac{t^{\mathrm{FD}}}{t^{\mathrm{FD}}_{\mathrm{Q}}} \tag{59a}$$

$$\text{s.t. } (1), (13), (17), (23), (58b), (58c),$$

$$R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)\frac{S_d}{R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)} + R_{2,k}(\boldsymbol{\zeta}_2)\frac{N_c D_{\max} c_{\max}}{f} + R^{\mathrm{FD}}_{3,k}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\frac{S_u}{R^{\mathrm{FD}}_u(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)} \ge t, \quad \forall k \tag{59b}$$

$$\frac{S_d}{R_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)} + \frac{N_c D_{\max} c_{\max}}{f} + \frac{S_u}{R^{\mathrm{FD}}_u(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)} \le t^{\mathrm{FD}}_{\mathrm{Q}}, \tag{59c}$$

$$t^{\mathrm{FD}}_{\mathrm{Q}} \le t^{\mathrm{FD}}_{\mathrm{QoS}}, \tag{59d}$$

where $\bar{\mathbf{x}}^{\mathrm{FD}} = [(\mathbf{x}^{\mathrm{FD}})^T, t^{\mathrm{FD}}, t^{\mathrm{FD}}_{\mathrm{Q}}]^T$, which is then equivalent to

$$\max_{\tilde{\mathbf{x}}^{\mathrm{FD}}} \; z^{\mathrm{FD}} \tag{60a}$$

$$\text{s.t. } (1), (13), (17), (23), (58b), (58c), (59d),$$

$$a_1 S_d + a_2 N_c D_{\max} c_{\max} + a^{\mathrm{FD}}_3 S_u \ge t^{\mathrm{FD}}, \quad \forall k \tag{60b}$$

$$\frac{S_d}{r_d} + \frac{N_c D_{\max} c_{\max}}{f} + \frac{S_u}{r^{\mathrm{FD}}_u} \le t^{\mathrm{FD}}_{\mathrm{Q}}, \tag{60c}$$

$$z^{\mathrm{FD}} t^{\mathrm{FD}}_{\mathrm{Q}} \le t^{\mathrm{FD}}, \tag{60d}$$

$$a_1 \tilde{r}_d \le r_{1,k}, \quad \forall k \tag{60e}$$

$$a_2 f \le r_{2,k}, \quad \forall k \tag{60f}$$

$$a^{\mathrm{FD}}_3 \tilde{r}^{\mathrm{FD}}_u \le r^{\mathrm{FD}}_{3,k}, \quad \forall k \tag{60g}$$

$$r_d \le R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall \ell \tag{60h}$$

$$r^{\mathrm{FD}}_u \le R^{\mathrm{FD}}_{u,\ell}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3), \quad \forall \ell \tag{60i}$$

$$r_{1,k} \le R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1), \quad \forall k \tag{60j}$$

$$r_{2,k} \le R_{2,k}(\boldsymbol{\zeta}_2), \quad \forall k \tag{60k}$$

$$r^{\mathrm{FD}}_{3,k} \le R^{\mathrm{FD}}_{3,k}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3), \quad \forall k \tag{60l}$$

$$R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) \le \tilde{r}_d, \quad \forall \ell \tag{60m}$$

$$R^{\mathrm{FD}}_{u,\ell}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \le \tilde{r}^{\mathrm{FD}}_u, \quad \forall \ell \tag{60n}$$

where $\tilde{\mathbf{x}}^{\mathrm{FD}} \triangleq [(\bar{\mathbf{x}}^{\mathrm{FD}})^T, r_d, r^{\mathrm{FD}}_u, a_1, a_2, a^{\mathrm{FD}}_3, \mathbf{r}_1^T, \mathbf{r}_2^T, (\mathbf{r}^{\mathrm{FD}}_3)^T, \tilde{r}_d, \tilde{r}^{\mathrm{FD}}_u, z^{\mathrm{FD}}]^T$, $\mathbf{r}_1 = \{r_{1,k}\}$, $\mathbf{r}_2 = \{r_{2,k}\}$, $\mathbf{r}^{\mathrm{FD}}_3 = \{r^{\mathrm{FD}}_{3,k}\}$. It is clear that the nonconvexity of problem (60) is due to (60d)-(60n). We remark that (60e) and (60f) are indeed (46b) and (46c), respectively, and their convex approximations are given in (53a) and (53b). Similar to (53c) and (53d), constraints (60d) and (60g) can be approximated by the following convex constraints

$$\frac{1}{4}\big[(z^{\mathrm{HD}} + t^{\mathrm{HD}}_{\mathrm{Q}})^2 - (z^{\mathrm{HD}(n)} - t^{\mathrm{HD}(n)}_{\mathrm{Q}})^2 + 2(z^{\mathrm{HD}} - t^{\mathrm{HD}}_{\mathrm{Q}}) \times (z^{\mathrm{HD}(n)} - t^{\mathrm{HD}(n)}_{\mathrm{Q}})\big] \le t^{\mathrm{HD}}, \tag{61a}$$

$$\frac{1}{4}\big[(a^{\mathrm{FD}}_3 + \tilde{r}^{\mathrm{FD}}_u)^2 - (a^{\mathrm{FD}(n)}_3 - \tilde{r}^{\mathrm{FD}(n)}_u)^2 + 2(a^{\mathrm{FD}}_3 - \tilde{r}^{\mathrm{FD}}_u) \times (a^{\mathrm{FD}(n)}_3 - \tilde{r}^{\mathrm{FD}(n)}_u)\big] \le r^{\mathrm{FD}}_{3,k}, \forall k \tag{61b}$$

Algorithm 2 Algorithm for Solving (58)

1: Input: Set $n = 0$ and choose an initial point $\tilde{\mathbf{x}}^{\text{FD}(0)} \in \tilde{\mathcal{F}}^{\text{FD}}$
2: **repeat**
3: Solve (63) to get $\tilde{\mathbf{x}}^{\text{FD}*}$
4: $\tilde{\mathbf{x}}^{\text{FD}(n+1)} \leftarrow \tilde{\mathbf{x}}^{\text{FD}*}$
5: $n \leftarrow n + 1$
6: **until** convergence

where $a_3^{\text{FD}(n)}$ and $\tilde{r}_u^{\text{FD}(n)}$ are the values of a_3^{FD} and $\tilde{r}_u^{\text{FD}}$ at the n-th iteration, respectively.

To proceed further, we note that the convex approximate constraints of (60h), (60j), (60k), and (60m) are already presented in (50a), (50c), (50d), and (55a), respectively. Also, (60i), (60l), and (60n) are similar to (44c), (45f), and (45c). Thus, following the same steps to obtaining (50b), (50e), and (55b), we can approximate (60i), (60l), and (60n) as the following convex constraints

$$r_u^{\text{FD}} \le \tilde{R}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3), \quad \forall \ell \tag{62a}$$

$$r_{3,k}^{\text{FD}} \le \tilde{R}_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3), \quad \forall k \tag{62b}$$

$$\hat{R}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \le \tilde{r}_u^{\text{FD}}, \quad \forall \ell \tag{62c}$$

where $\tilde{R}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)$, $\tilde{R}_{3,k}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)$, and $\hat{R}_{u,\ell}^{\text{FD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)$ are given in (77) and (78) in Appendix C.

At iteration $n + 1$, for a given point $\tilde{\mathbf{x}}^{\text{FD}(n)}$, problem (58) is approximated by the following convex problem:

$$\max \{z^{\text{FD}} \mid \tilde{\mathbf{x}}^{\text{FD}} \in \tilde{\mathcal{F}}^{\text{FD}}\}, \tag{63}$$

where $\tilde{\mathcal{F}}^{\text{FD}} \triangleq \{(1), (13), (17), (23), (58b), (58c), (59d), (60b), (60c), (53a), (53b), (50a), (50c), (50d), (55a), (61), (62)\}$. We outline the main steps to solve problem (60) in Algorithm 2.

Remark 2: Similar to Algorithm 1, Algorithm 2 requires a feasible point to (60) which is not trivial for find, especially when the SI is high. To overcome this issue we follow the same procedure as described in Remark 1. Specifically, if scaling randomly generated variables cannot produce a feasible solution, we introduce add a slack variable s and consider the following problem

$$\max_{s \le 0, \tilde{\mathbf{x}}^{\text{FD}}} \; z^{\text{FD}} + s \tag{64a}$$

$$\text{s.t. } \tilde{\mathbf{x}}^{\text{FD}} \in \tilde{\mathcal{F}}^{\text{FD}} \setminus (59d), \tag{64b}$$

$$t_{\text{Q}}^{\text{FD}} + s \le t_{\text{QoS}}^{\text{FD}}. \tag{64c}$$

Then, problem (64) is solved iteratively until convergence. If $|s|$ is smaller a pre-determined error tolerance, we will take $\tilde{\mathbf{x}}^{\text{FD}}$ as the final solution. Otherwise, (60) (and thus (58)) is said to be infeasible.

Remark 3: In practical wireless networks, channel estimation is a critical part. In the previous section, we have explained how channels can be estimated in all steps of the proposed system model. Specifically, users send pilot signals to the BS to estimate the channels. We note that the channels are estimated at the BS, not at users. We also note that channel estimation is only required for the BS for beamforming in the downlink channel, and the achievable rate derived in the paper is based on the assumption that channel estimations are not available in the downlink channel since no pilots are sent from the BS. This is a standard approach in massive MIMO using TDD. We remark that in our paper, optimization problems are solved at the BS, and thus, the users do not need to obtain channel information or spend any computing resources. In other words, we adopt a centralized optimization method in this paper. After solving the optimization problems, the BS sends the power coefficients to the UEs through dedicated feedback channels. Importantly, the optimization problems are performed on the large-scale fading time scale which changes slowly with time.

IV. Numerical Results

A. Parameter Setting

We consider a $D \times D$ m^2 area where the BS is at the centre, while L FL UEs and K non-FL UEs are randomly distributed. The large-scale fading coefficients are modeled in the same manner as [35, Eq. (46)]:

$$\beta_\ell[\text{dB}] = -148.1 - 37.6 \log_{10}\left(\frac{d_\ell}{1 \text{ km}}\right) + z_\ell, \tag{65}$$

where $d_\ell \ge 35$ m is the distance between UE ℓ and the BS, z_ℓ is a shadow fading coefficient which is modeled using a log-normal distribution having zero mean and 7 dB standard deviation. We set $N_0 = -92$ dBm, $t_{\text{QoS}} = 3$ s, $B = 20$ MHz, $\rho_d = 10$ W, $\rho_u = \rho_p = 0.2$ W, $\tau_{d,p} = \tau_{u,p} = 20$, $\tau_{S_1,p} = \tau_{S_2,p} = \tau_{S_3,p} = 20$, $\tau_c = 200$, $f_{\min} = 0$, $f_{\max} = 5 \times 10^9$ cycles/s, $D_\ell = D_{\max} = 1.6 \times 10^5$ samples, $c_\ell = c_{\max} = 20$ cycles/sample, $N_c = 20$, $S_d = S_u = 16 \times 10^6$ bits or 16Mb. The path loss β_{SI} is taken as $\beta_{\text{SI}} = 10^{\frac{\text{PL}}{10}}$, where PL $= -81.1846$ dB [36]. If not otherwise mentioned, the value of $\sigma_{\text{SI},0}^2/N_0$ is set to 20 dB. In Fig. 2, we plot the convergence behavior of proposed algorithms for two random channel realizations. In the remaining figures, the minimum effective rate of non-FL UEs is plotted by averaging it over 100 random channel realizations. Before proceeding further, we remark that there could be considerable amount of FL UEs present compared to the non-FL UEs in many applications of future wireless networks due to the growing interest of machine learning techniques in wireless communication. Thus, the number of FL and non-FL UEs are assumed to be equal for simulations purpose, except where otherwise stated.

B. Results and Discussions

Since we are the first to introduce a system model to support both groups of FL and non-FL users, we are not aware of any existing closely related baseline schemes to benchmark our proposed solutions in this paper. Therefore, we ourselves present two baseline schemes based on equal power allocation (EPA) and frequency-division multiple access (FDMA) methods, and compare them to the proposed solutions. We remark that the EPA and FDMA are well-known approaches in wireless communications, and we merely customize them to fit the considered system model. The purpose is to show the potential gains of optimizing the involving parameters, compared to the two simple heuristic solutions. As an introductory work, we believe that comparing our proposed solutions with these baseline schemes is sufficient enough. The two considered baseline schemes are detailed as follows.

- **BL1**: Steps (S1) and (S3) of this scheme have the same designs as shown in the proposed scheme. In Step

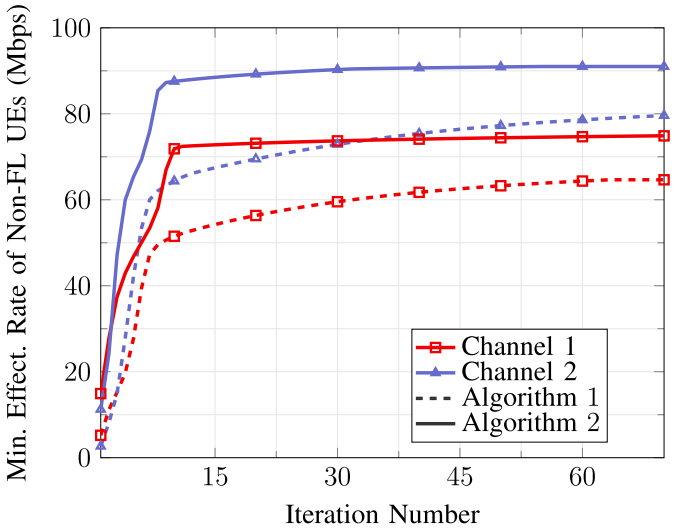


Fig. 2. Convergence of the proposed Algorithm 1 and Algorithm 2 for two different channel realizations. Here $L = K = 5$, and $M = 50$.

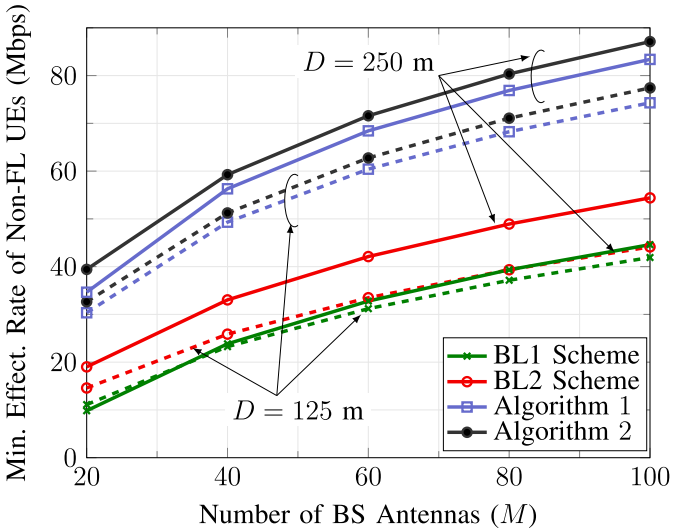


Fig. 3. Minimum Effective rate of non-FL UEs for different values of number of BS antennas. Here $L = K = 5$.

(S3), the uplink transmission for the FL group and the downlink of the non-FL group are executed using a frequency-division multiple access (FDMA) approach for transmission. In particular, we equally allocate the frequency band to all UEs such that each FL UE or non-FL UE has one single bandwidth subband for its transmission. This FDMA scheme is widely used in FL literature (e.g., [20], [37]). The uplink and downlink rates of FL UEs in **BL1** are derived in Appendix B. The optimization problem of **BL1** has the same mathematical structure as that of the proposed scheme. Therefore, it can be solved by slightly modifying Algorithm 1 using the same approximations.

- **BL2**: The downlink powers to FL and non-FL UEs in Step (S1) are equal, i.e., $\eta_{d,\ell} = \zeta_{1,k} = \frac{1}{L+K}, \forall \ell, k$. The downlink powers to non-FL UEs in Step (S2) and (S3) are also the same, i.e, $\zeta_{2,k} = \zeta_{3,k} = \frac{1}{K}, \forall k$. In addition, in Step (S3), each FL UE uses full power, i.e, $\eta_{u,\ell} = 1, \forall \ell$. The processing frequencies are $f = \frac{N_c D_m c_m}{t_{\text{QoS}} - t_d - t_u}$.

We first provide the convergence of the proposed scheme in comparison with BL1 and BL2 schemes. The convergence plot is shown in Fig. 2. It can be observed that both algorithms converge in less than 30 iterations for both channel realizations. Further, we note that for both channels, FD-based solution provides a better objective than the HD-based solution when SI is 20 dB as considered in this figure.

Next, in Figs. 3 and 4, we compare the minimum effective rate of the non-FL UEs by the two proposed schemes and the two considered baseline schemes. As seen clearly, both proposed schemes offer a better performance than the baseline counterparts. The figures not only demonstrate the significant advantage of a joint allocation of power and computing frequency over the heuristic scheme **BL2**, but also show the benefit of using massive MIMO. Thanks to massive MIMO technology, the data rate of each non-FL UE increases when the number of antennas increases, which then leads to a significant increase in the minimum effective data rates.

Moreover, Figs. 3 and 4 also confirm that in each frequency band used for each group, serving all the UEs simultaneously is better than serving them using the EPA approach. Specifically, the proposed approaches outperform **BL2** in almost

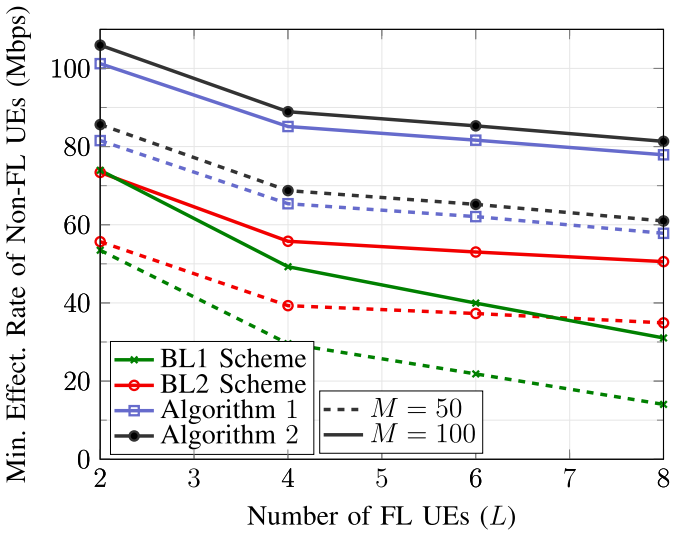


Fig. 4. Minimum Effective rate of non-FL UEs for different values of number of FL UEs. Here $K = 5$, and $M = 50$.

every case. The gap between the proposed schemes and **BL1** is even bigger when the number of FL UEs increases. This is because the effective rate of non-FL UE k can be considered as the weighted rate $R_k \triangleq \frac{R_{1,k} t_d + R_{2,k} t_C + R_{3,k} t_u}{t_d + t_C + t_u}$, where t_d, t_C, t_u are the weights associated with $R_{1,k}$, $R_{2,k}$, and $R_{3,k}$, respectively. Here, $R_{2,k}$ is the dominant term because in Step (S2), all the non-FL UEs are served simultaneously without interference from FL UEs. In **BL1**, $R_{u,\ell}$ is very small due to its prelog factor $\frac{1}{L+K}$, which leads to a large t_u. When the weight t_u becomes dominant compared to t_d and t_C, R_k of **BL1** is close to $R_{u,\ell}$ which is much lower than R_k of the proposed schemes. As L increases, $R_{u,\ell}$ decreases further and hence, R_k also decreases.

We now investigate the effect of high SI on the performance of the FD-based solution to understand when the FD-based algorithm is superior to the HD counterpart. For this purpose, the minimum effective rate of non-FL UEs is plotted in Fig. 5 for different values of $\sigma^2_{\text{SI},0}/N_0$. We also introduce a hybrid scheme which selects the approach that has the better objective among the two. For low values of SI (i.e., upto 65 dB), the FD-based approach performs better, which is expected and thus, the hybrid scheme is the same as the FD-based scheme.

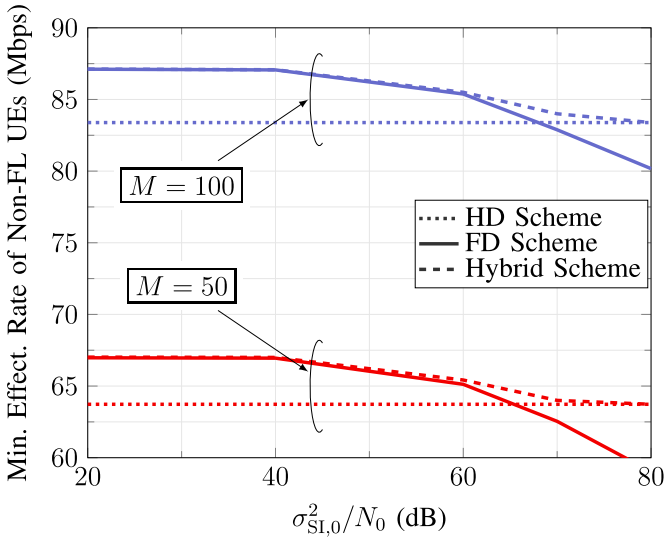
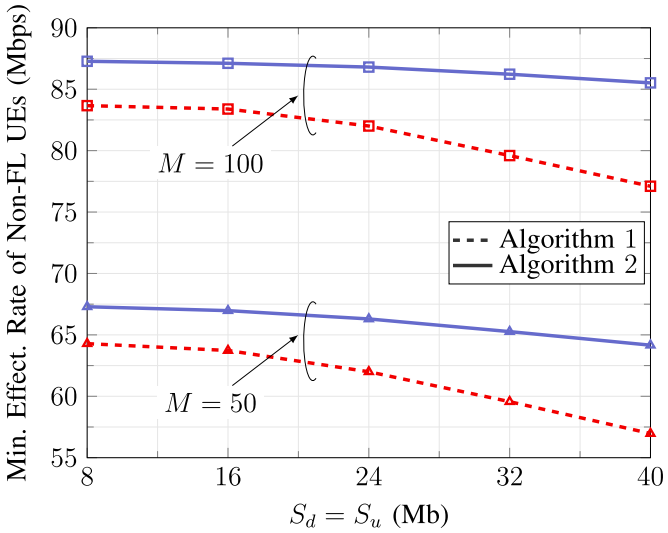
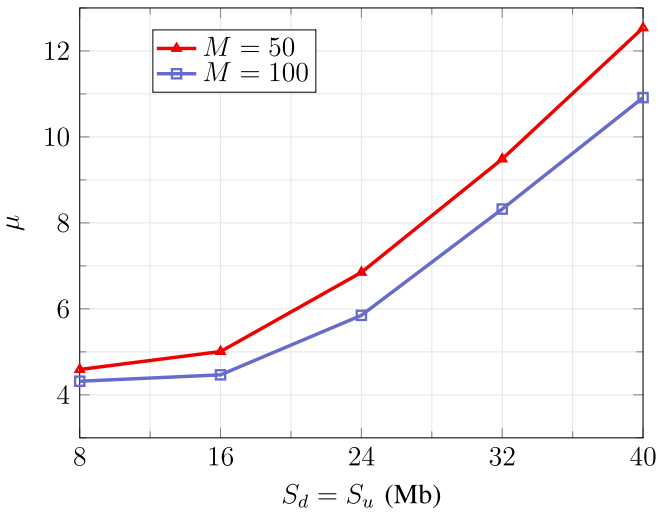


Fig. 5. Minimum Effective Rate of non-FL UEs for different values of $\sigma^2_{\mathrm{SI},0}/N_0$ in dB.



(a) Min. Effect. Rate of Non-FL UEs VS $S_d = S_u$



(b) Percentage advantage VS $S_d = S_u$

Fig. 6. Minimum effective rate of non-FL UEs for different values of S_d and S_u. Here $L = K = 5$.

On the other hand, for large values of SI (i.e., beyond 65 dB), the effectiveness of the FD-based approach starts to decrease due to the increased SI between the FL and non-FL groups. Especially, the HD-based scheme outperforms the FD-based approach when the SI is around than 80 dB. Thus, the hybrid scheme is equal to the HD-based scheme for very large SI as can be seen clearly in Fig. 5.

In the final numerical experiment, we plot the minimum effective rate of non-FL UEs against the values of S_d and S_u in Fig. 6 to compare the performance of both proposed algorithms for different data sizes. We introduce a parameter $\mu \triangleq \frac{R_k^{\mathrm{FD}} - R_k^{\mathrm{HD}}}{R_k^{\mathrm{HD}}} \times 100$, which defines the gain in percentage of the FD-based solution over the HD-based solution. From Fig. 6, we can observe that as the data size increases, the performance of FD-based scheme decreases very slowly compared to the HD-based scheme. As a result, the gain in percentage of the FD-based scheme over the HD-based scheme, μ, increases with the data size. Thus, we can conclude that for problems with large sizes of FL model updates, the FD-based scheme should be preferred over the HD-based scheme for SI of 20 dB.

V. Conclusion

We have presented two communication schemes that can support both the FL and non-FL UEs using massive MIMO technology which has not been considered previously. In particular, we have defined and maximize the effective rate of downlink non-FL UEs in presence of a QoS latency constraint on FL UEs. We have also presented the HD and FD based solutions to the considered problem, specifically during the uplink of each FL iteration where the uplink FL UEs and downlink non-FL UEs receive their corresponding data simultaneously. In the downlink of each FL iteration, both FL and non-FL UEs continue to be served in the same time-frequency resource. The simulation results have showed that the proposed HD-based and FD-based schemes outperform the considered baseline schemes in all considered scenarios. It has also been shown that the FD-based scheme is superior to the HD-based scheme for the SI of upto 70 dB. The FD-based scheme is also more beneficial than the HD-based scheme in terms of the effective rate achieved by the non-FL UEs in the cases of the model updates of large sizes.

Appendix A
Achievable Rates for FD in Step (S3)

Uplink Transmission of FL UEs: In this appendix, we simplify (31). Particularly, we need to find two variance terms: $\mathbb{V}\mathrm{ar}\{\mathbf{u}_{u,\ell}^H \mathbf{E}_u \mathbf{D}_{\boldsymbol{\eta}_u}^{1/2} \mathbf{s}_u\}$ and $\mathbb{V}\mathrm{ar}\{\mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2} \mathbf{s}_u\}$. For the first term, we know that $\mathbf{Z}_u$ is independent of $\mathbf{E}_u$. Thus, we can state that

$$\mathbb{V}\mathrm{ar}\{\mathbf{u}_{u,\ell}^H \mathbf{E}_u \mathbf{D}_{\boldsymbol{\eta}_u}^{1/2} \mathbf{s}_u\} = \sum_{i \in \mathcal{L}} (\beta_i - \sigma_{u,i}^2)\eta_{u,i}. \tag{66}$$

$$\begin{aligned}
&\mathbb{V}\mathrm{ar}\{\mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2} \mathbf{s}_u\} \\
&= \mathbb{E}\{|\, \mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2} \mathbf{s}_u \,|^2\} \\
&= \mathbb{E}\{\mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2} \mathbf{s}_u \mathbf{s}_u^H \mathbf{D}_{\boldsymbol{\zeta}_3}^{1/2} \mathbf{U}_3^H \mathbf{G}_{\mathrm{SI}}^H \mathbf{u}_{u,\ell}\} \\
&= \mathbb{E}\{\mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{D}_{\boldsymbol{\zeta}_3} \mathbf{U}_3^H \mathbf{G}_{\mathrm{SI}}^H \mathbf{u}_{u,\ell}\} \\
&= (\sum_{i \in \mathcal{K}} \zeta_{3,i})\, \mathbb{E}\{\mathbf{u}_{u,\ell}^H \mathbf{G}_{\mathrm{SI}} \mathbf{U}_3 \mathbf{U}_3^H \mathbf{G}_{\mathrm{SI}}^H \mathbf{u}_{u,\ell}\}
\end{aligned}$$

$$= (\sum_{i\in\mathcal{K}} \zeta_{3,i})\, \mathbb{E}\{\mathbf{u}_{u,\ell}^H\, \mathbf{G}_{\text{SI}}\, \mathbf{G}_{\text{SI}}^H\, \mathbf{u}_{u,\ell}\}. \tag{67}$$

Using law of large numbers, $\mathbb{E}\{\mathbf{G}_{\text{SI}}\, \mathbf{G}_{\text{SI}}^H\} \approx M\beta_{\text{SI}}\sigma_{\text{SI},0}^2$. Therefore, the above equation can be approximated as

$$\begin{aligned}\mathbb{V}\text{ar}\{\mathbf{u}_{u,\ell}^H\, \mathbf{E}_u\, \mathbf{D}_{\boldsymbol{\eta}_u}^{1/2}\, \mathbf{s}_u\} &\approx M\beta_{\text{SI}}\sigma_{\text{SI},0}^2(\sum_{i\in\mathcal{K}} \zeta_{3,i})\, \mathbb{E}\{\mathbf{u}_{u,\ell}^H\, \mathbf{u}_{u,\ell}\} \\ &= M\beta_{\text{SI}}\sigma_{\text{SI},0}^2 \sum_{i\in\mathcal{K}} \zeta_{3,i}. \end{aligned}\tag{68}$$

Appendix B
Achievable Rates and Data Received for Baseline 1 (BL1) Scheme

For the FD communication in Step (S3), FL UEs observe SI from non-FL UEs, while non-FL UEs observe IGI from FL UEs. Assume the FDMA approach is used in this scheme. The entire system bandwidth for FL UEs and non-FL UEs are divided into $(L+K)$ subbands, and in each bandwidth subband, only one FL or non-FL user is allowed to transmit, and thus, there is no interference between any other FL and non-FL user. In this context, the FDMA approach is the same for both HD scheme and FD scheme. The uplink rate for the FL group is expressed as

$$R_u^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) \triangleq \min_{\ell\in\mathcal{L}} \frac{\tau_c - \tau_{u,p}}{(L+K)\tau_c} B \log_2\left(1 + \text{SINR}_{u,\ell}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right), \tag{69}$$

where $\tau_{u,p} = 1$ and $\text{SINR}_{u,\ell}^{\text{FDMA}}(\boldsymbol{\eta}_u)$ is given by

$$\text{SINR}_{u,\ell}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_u \eta_{u,\ell} M \sigma_{u,\ell}^2}{1 + \rho_u \beta_\ell \eta_{u,\ell}}. \tag{70}$$

Then, the time for each FL UE to complete its uplink transmission is the same, which is given by

$$t_u^{\text{FDMA}}(\boldsymbol{\eta}_u) = \frac{S_u}{R_u^{\text{FDMA}}(\boldsymbol{\eta}_u)}. \tag{71}$$

Similarly for non-FL UEs, the downlink rate is written as

$$R_{3,k}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\tau_c - \tau_{3,p}}{(L+K)\tau_c} B \log_2\left(1 + \text{SINR}_{3,k}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right), \tag{72}$$

where $\tau_{3,p} = 1$ and $\text{SINR}_{3,k}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)$ is calculated as

$$\text{SINR}_{3,k}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\rho_d \zeta_{3,k} M \sigma_{3,k}^2}{1 + \rho_d \bar{\beta}_k \zeta_{3,k}}. \tag{73}$$

The transmission time from each FL UE to the BS and the amount of downlink data received at all non-FL UE $k, \forall k \in \mathcal{K}$ are given by (27) and (28), respectively. The optimization problem in this scheme is

$$\max_{\mathbf{x}^{\text{FDMA}}} \frac{\min_{k\in\mathcal{K}} \left(D_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + D_{2,k}(f, \boldsymbol{\zeta}_2) + D_{3,k}^{\text{FDMA}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3)\right)}{t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t_u^{\text{FMDA}}(\boldsymbol{\eta}_u)} \tag{74a}$$

$$\text{s.t. } (1), (13), (17), (23), (41b), (41c),$$
$$t_d(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) + t_C(f) + t_u^{\text{FDMA}}(\boldsymbol{\eta}_u) \le t_{\text{QoS}}. \tag{74b}$$

Appendix C
Expressions of Lower and Upper Bounds of Rate Functions

From (48), the concave lower bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u)$, $R_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$, $R_{2,k}(\boldsymbol{\zeta}_2)$, and $R_{3,k}^{\text{HD}}(\boldsymbol{\zeta}_3)$ are found as given in (75a)-(75e), as shown at the bottom of the page, where $\psi_{d,\ell} = \rho_d(M-L-K)\sigma_{d,\ell}^2\eta_{d,\ell}, \psi_{d,\ell}^{(n)} = \rho_d(M-L-K)\sigma_{d,\ell}^2\eta_{d,\ell}^{(n)}, \theta_{d,\ell} = 1 + \rho_d(\beta_\ell - \sigma_{d,\ell}^2)\sum_{i\in\mathcal{L}}\eta_{d,i} + \rho_d\beta_\ell\sum_{k\in\mathcal{K}}\zeta_{1,k}, \theta_{d,\ell}^{(n)} = 1 + \rho_d(\beta_\ell - \sigma_{d,\ell}^2)\sum_{i\in\mathcal{L}}\eta_{d,i}^{(n)} + \rho_d\beta_\ell\sum_{k\in\mathcal{K}}\zeta_{1,k}^{(n)}, \psi_{u,\ell} = \rho_u(M-L)\sigma_{u,\ell}^2\eta_{u,\ell}, \psi_{u,\ell}^{(n)} = \rho_u(M-L)\sigma_{u,\ell}^2\eta_{u,\ell}^{(n)}, \theta_{u,\ell}^{\text{HD}} = 1 + \rho_u\sum_{i\in\mathcal{L}}(\beta_i - \sigma_{u,i}^2)\eta_{u,i}, \theta_{u,\ell}^{\text{HD}(n)} = 1 + \rho_u\sum_{i\in\mathcal{L}}(\beta_i - \sigma_{u,i}^2)\eta_{u,i}^{(n)}, \psi_{1,k} = \rho_d(M-L-K)\sigma_{1,k}^2\zeta_{1,k}, \psi_{1,k}^{(n)} = \rho_d(M-L-K)\sigma_{1,k}^2\zeta_{1,k}^{(n)}, \theta_{1,k} = 1 + \rho_d(\bar{\beta}_k - \sigma_{1,k}^2)\sum_{i\in\mathcal{K}}\zeta_{1,i} + \rho_d\bar{\beta}_k\sum_{\ell\in\mathcal{L}}\eta_{d,\ell}, \theta_{1,k}^{(n)} = 1 + \rho_d(\bar{\beta}_k - \sigma_{1,k}^2)\sum_{i\in\mathcal{K}}\zeta_{1,i}^{(n)} + \rho_d\bar{\beta}_k\sum_{\ell\in\mathcal{L}}\eta_{d,\ell}^{(n)}; \psi_{2,k} = \rho_d(M-K)\sigma_{1,k}^2\zeta_{2,k}, \psi_{2,k}^{(n)} = \rho_d(M-K)\sigma_k^2\zeta_{2,k}^{(n)}, \theta_{2,k} = 1 + \rho_d(\bar{\beta}_k - \sigma_{2,k}^2)\sum_{i\in\mathcal{K}}\zeta_{2,i}, \theta_{2,k}^{(n)} = 1 + \rho_d(\bar{\beta}_k - \sigma_{2,k}^2)\sum_{i\in\mathcal{K}}\zeta_{2,i}^{(n)}, \psi_{3,k} = \rho_d(M-K)\sigma_{3,k}^2\zeta_{3,k}, \psi_{3,k}^{(n)} = \rho_d(M-K)\sigma_{3,k}^2\zeta_{3,k}^{(n)}, \theta_{3,k}^{\text{HD}} = 1 + \rho_d(\bar{\beta}_k - \sigma_{3,k}^2)\sum_{i\in\mathcal{K}}\zeta_{3,i}, \theta_{3,k}^{\text{HD}(n)} = 1 + \rho_d(\bar{\beta}_k - \sigma_{3,k}^2)\sum_{i\in\mathcal{K}}\zeta_{3,i}^{(n)}$.

From (54) convex upper bounds of $R_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1)$ and $R_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u)$ are given in (76), as shown at the top of the next page.

$$\tilde{R}_{d,\ell}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) = \frac{\tau_c - \tau_{d,p}}{\tau_c \log 2} B\Big[\log\left(1 + \frac{\psi_{d,\ell}^{(n)}}{\theta_{d,\ell}^{(n)}}\right) + \frac{2\,\psi_{d,\ell}^{(n)}}{\psi_{d,\ell}^{(n)} + \theta_{d,\ell}^{(n)}} - \frac{(\psi_{d,\ell}^{(n)})^2}{(\psi_{d,\ell}^{(n)} + \theta_{d,\ell}^{(n)})\,\psi_{d,\ell}} - \frac{\psi_{d,\ell}^{(n)}\,\theta_{d,\ell}}{(\psi_{d,\ell}^{(n)} + \theta_{d,\ell}^{(n)})\,\theta_{d,\ell}^{(n)}}\Big] \tag{75a}$$

$$\tilde{R}_{u,\ell}^{\text{HD}}(\boldsymbol{\eta}_u) = \frac{\tau_c - \tau_{u,p}}{\tau_c \log 2} \frac{B}{2}\Big[\log\left(1 + \frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\text{HD}(n)}}\right) + \frac{2\,\psi_{u,\ell}^{(n)}}{\psi_{u,\ell}^{(n)} + \theta_{u,\ell}^{\text{HD}(n)}} - \frac{(\psi_{u,\ell}^{(n)})^2}{(\psi_{u,\ell}^{(n)} + \theta_{u,\ell}^{\text{HD}(n)})\,\psi_{u,\ell}} - \frac{\psi_{u,\ell}^{(n)}\,\theta_{u,\ell}^{\text{HD}}}{(\psi_{u,\ell}^{(n)} + \theta_{u,\ell}^{\text{HD}(n)})\,\theta_{u,\ell}^{\text{HD}(n)}}\Big] \tag{75b}$$

$$\tilde{R}_{1,k}(\boldsymbol{\eta}_d, \boldsymbol{\zeta}_1) = \frac{\tau_c - \tau_{1,p}}{\tau_c \log 2} B\Big[\log\left(1 + \frac{\psi_{1,k}^{(n)}}{\theta_{1,k}^{(n)}}\right) + \frac{2\,\psi_{1,k}^{(n)}}{\psi_{1,k}^{(n)} + \theta_{1,k}^{(n)}} - \frac{(\psi_{1,k}^{(n)})^2}{(\psi_{1,k}^{(n)} + \theta_{1,k}^{(n)})\,\psi_{1,k}} - \frac{\psi_{1,k}^{(n)}\,\theta_{1,k}}{(\psi_{1,k}^{(n)} + \theta_{1,k}^{(n)})\,\theta_{1,k}^{(n)}}\Big] \tag{75c}$$

$$\tilde{R}_{2,k}(\boldsymbol{\zeta}_2) = \frac{\tau_c - \tau_{2,p}}{\tau_c \log 2} B\Big[\log\left(1 + \frac{\psi_{2,k}^{(n)}}{\theta_{2,k}^{(n)}}\right) + \frac{2\,\psi_{2,k}^{(n)}}{\psi_{2,k}^{(n)} + \theta_{2,k}^{(n)}} - \frac{(\psi_{2,k}^{(n)})^2}{(\psi_{2,k}^{(n)} + \theta_{2,k}^{(n)})\,\psi_{2,k}} - \frac{\psi_{2,k}^{(n)}\,\theta_{2,k}}{(\psi_{2,k}^{(n)} + \theta_{2,k}^{(n)})\,\theta_{2,k}^{(n)}}\Big] \tag{75d}$$

$$\tilde{R}_{3,k}^{\text{HD}}(\boldsymbol{\eta}_u, \boldsymbol{\zeta}_3) = \frac{\tau_c - \tau_{3,p}}{\tau_c \log 2} \frac{B}{2}\Big[\log\left(1 + \frac{\psi_{3,k}^{(n)}}{\theta_{3,k}^{\text{HD}(n)}}\right) + \frac{2\,\psi_{3,k}^{(n)}}{\psi_{3,k}^{(n)} + \theta_{3,k}^{\text{HD}(n)}} - \frac{(\psi_{3,k}^{(n)})^2}{(\psi_{3,k}^{(n)} + \theta_{3,k}^{\text{HD}(n)})\,\psi_{3,k}} - \frac{\psi_{3,k}^{(n)}\,\theta_{3,k}^{\text{HD}}}{(\psi_{3,k}^{(n)} + \theta_{3,k}^{\text{HD}(n)})\,\theta_{3,k}^{\text{HD}(n)}}\Big] \tag{75e}$$

$$\hat{R}_{d,\ell}(\boldsymbol{\eta}_d,\boldsymbol{\zeta}_1)=\frac{\tau_c-\tau_{d,p}}{\tau_c\log 2}B\Big[\log\big(1+\frac{\psi_{d,\ell}^{(n)}}{\theta_{d,\ell}^{(n)}}\big)+\frac{\theta_{d,\ell}^{(n)}}{\psi_{d,\ell}^{(n)}+\theta_{d,\ell}^{(n)}}-\frac{(\psi_{d,\ell})^2+(\psi_{d,\ell}^{(n)})^2}{2\times\psi_{d,\ell}^{(n)}\times\theta_{d,\ell}}-\frac{\psi_{d,\ell}^{(n)}}{\theta_{d,\ell}^{(n)}}\Big], \tag{76a}$$

$$\hat{R}_{u,\ell}^{\mathrm{HD}}(\boldsymbol{\eta}_u)=\frac{\tau_c-\tau_{u,p}}{\tau_c\log 2}\frac{B}{2}\Big[\log\big(1+\frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\mathrm{HD}(n)}}\big)+\frac{\theta_{u,\ell}^{\mathrm{HD}(n)}}{\psi_{u,\ell}^{(n)}+\theta_{u,\ell}^{\mathrm{HD}(n)}}-\frac{(\psi_{u,\ell})^2+(\psi_{u,\ell}^{(n)})^2}{2\,\psi_{u,\ell}^{(n)}\,\theta_{u,\ell}^{\mathrm{HD}}}-\frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\mathrm{HD}(n)}}\Big]. \tag{76b}$$

$$\tilde{R}_{u,\ell}^{\mathrm{FD}}(\boldsymbol{\eta}_u)=\frac{\tau_c-\tau_{u,p}}{\tau_c\log 2}\frac{B}{2}\Big[\log\big(1+\frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\mathrm{FD}(n)}}\big)+\frac{2\,\psi_{u,\ell}^{(n)}}{\psi_{u,\ell}^{(n)}+\theta_{u,\ell}^{\mathrm{FD}(n)}}-\frac{(\psi_{u,\ell}^{(n)})^2}{(\psi_{u,\ell}^{(n)}+\theta_{u,\ell}^{\mathrm{FD}(n)})\,\psi_{u,\ell}}-\frac{\psi_{u,\ell}^{(n)}\,\theta_{u,\ell}^{\mathrm{FD}}}{(\psi_{u,\ell}^{(n)}+\theta_{u,\ell}^{\mathrm{FD}(n)})\,\theta_{u,\ell}^{\mathrm{FD}(n)}}\Big], \tag{77a}$$

$$\tilde{R}_{3,k}^{\mathrm{FD}}(\boldsymbol{\eta}_u,\boldsymbol{\zeta}_3)=\frac{\tau_c-\tau_{3,p}}{\tau_c\log 2}\frac{B}{2}\Big[\log\big(1+\frac{\psi_{3,k}^{(n)}}{\theta_{3,k}^{\mathrm{FD}(n)}}\big)+\frac{2\,\psi_{3,k}^{(n)}}{\psi_{3,k}^{(n)}+\theta_{3,k}^{\mathrm{FD}(n)}}-\frac{(\psi_{3,k}^{(n)})^2}{(\psi_{3,k}^{(n)}+\theta_{3,k}^{\mathrm{FD}(n)})\,\psi_{3,k}}-\frac{\psi_{3,k}^{(n)}\,\theta_{3,k}^{\mathrm{FD}}}{(\psi_{3,k}^{(n)}+\theta_{3,k}^{\mathrm{FD}(n)})\,\theta_{3,k}^{\mathrm{FD}(n)}}\Big], \tag{77b}$$

$$\hat{R}_{u,\ell}^{\mathrm{FD}}(\boldsymbol{\eta}_u)=\frac{\tau_c-\tau_{u,p}}{\tau_c\log 2}\frac{B}{2}\Big[\log\big(1+\frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\mathrm{FD}(n)}}\big)+\frac{\theta_{u,\ell}^{\mathrm{FD}(n)}}{\psi_{u,\ell}^{(n)}+\theta_{u,\ell}^{\mathrm{FD}(n)}}-\frac{(\psi_{u,\ell})^2+(\psi_{u,\ell}^{(n)})^2}{2\,\psi_{u,\ell}^{(n)}\,\theta_{u,\ell}^{\mathrm{FD}}}-\frac{\psi_{u,\ell}^{(n)}}{\theta_{u,\ell}^{\mathrm{FD}(n)}}\Big], \tag{78}$$

Similarly, the concave lower bounds of $R_{u,\ell}^{\mathrm{FD}}(\boldsymbol{\eta}_u,\boldsymbol{\zeta}_3)$ and $R_{3,k}^{\mathrm{FD}}(\boldsymbol{\eta}_u,\boldsymbol{\zeta}_3)$ for the FD scheme given in (77), as shown at the top of the page, and the convex upper bound of $R_{u,\ell}^{\mathrm{FD}}(\boldsymbol{\eta}_u)$ is given in (78), as shown at the top of the page, where $\theta_{u,\ell}^{\mathrm{FD}}=1+\rho_u\sum_{i\in\mathcal{L}}(\beta_i-\sigma_{u,i}^2)\eta_{u,i}+\rho_d M\beta_{\mathrm{SI}}\sigma_{\mathrm{SI},0}^2\sum_{j\in\mathcal{K}}\zeta_{3,j}, \theta_{u,\ell}^{\mathrm{FD}(n)}=1+\rho_u\sum_{i\in\mathcal{L}}(\beta_i-\sigma_{u,i}^2)\eta_{u,i}^{(n)}+\rho_d M\beta_{\mathrm{SI}}\sigma_{\mathrm{SI},0}^2\sum_{j\in\mathcal{K}}\zeta_{3,j}^{(n)}, \theta_{3,k}^{\mathrm{FD}}=1+\rho_d(\bar{\beta}_k-\sigma_{3,k}^2)\sum_{i\in\mathcal{K}}\zeta_{3,i}+\rho_u\sum_{i\in\mathcal{L}}\beta_{\mathrm{IGI},ki}\eta_{u,i}, \theta_{3,k}^{\mathrm{FD}(n)}=1+\rho_d(\bar{\beta}_k-\sigma_{3,k}^2)\sum_{i\in\mathcal{K}}\zeta_{3,i}^{(n)}+\rho_u\sum_{i\in\mathcal{L}}\beta_{\mathrm{IGI},ki}\eta_{u,i}^{(n)}$.

REFERENCES

[1] M. Farooq, T. T. Vu, H. Quoc Ngo, and L. Tran, "Serving federated learning and non-federated learning users: A massive MIMO approach," in *Proc. IEEE 23rd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jul. 2022, pp. 1–5.

[2] A. Ometov et al., "A survey on wearable technology: History, state-of-the-art and current challenges," *Comput. Netw.*, vol. 193, Jul. 2021, Art. no. 108074.

[3] T. Giannetsos, T. Dimitriou, and N. R. Prasad, "People-centric sensing in assistive healthcare: Privacy challenges and directions," *Secur. Commun. Netw.*, vol. 4, no. 11, pp. 1295–1307, Nov. 2011.

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

[5] E. Ahmed et al., "Bringing computation closer toward the user network: Is edge computing the solution?" *IEEE Commun. Mag.*, vol. 55, no. 11, pp. 138–144, Nov. 2017.

[6] B. McMahan, E. Moore, D. Ramage, S. Hampson, and B. A. Y. Arcas, "Communication-efficient learning of deep networks from decentralized data," in *Proc. Int. Conf. Artif. Intell. Stat. (AISTATS)*, 2017, pp. 1273–1282.

[7] I. S. Rubinstein and N. Good, "Privacy by design: A counterfactual analysis of Google and Facebook privacy incidents," *Berkeley Tech. Law J.*, vol. 28, no. 2, p. 1333, Jan. 2023.

[8] *We are Making on-Device AI Ubiquitous*, Qualcomm, San Diego, CA, USA, 2018.

[9] S. Abdulrahman, H. Tout, H. Ould-Slimane, A. Mourad, C. Talhi, and M. Guizani, "A survey on federated learning: The journey from centralized to distributed on-site learning and beyond," *IEEE Internet Things J.*, vol. 8, no. 7, pp. 5476–5497, Apr. 2021.

[10] M. Aledhari, R. Razzak, R. M. Parizi, and F. Saeed, "Federated learning: A survey on enabling technologies, protocols, and applications," *IEEE Access*, vol. 8, pp. 140699–140725, 2020.

[11] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, "FedHealth: A federated transfer learning framework for wearable healthcare," *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, Jul. 2020.

[12] W. Yang, N. Wang, Z. Guan, L. Wu, X. Du, and M. Guizani, "A practical cross-device federated learning framework over 5G networks," *IEEE Wireless Commun.*, vol. 29, no. 6, pp. 128–134, Dec. 2022.

[13] D. Shi, L. Li, R. Chen, P. Prakash, M. Pan, and Y. Fang, "Toward energy-efficient federated learning over 5G+ mobile devices," *IEEE Wireless Commun.*, vol. 29, no. 5, pp. 44–51, Oct. 2022.

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

[15] M. Chen, Z. Yang, W. Saad, C. Yin, H. V. Poor, and S. Cui, "A joint learning and communications framework for federated learning over wireless networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 269–283, Jan. 2021.

[16] M. M. Amiri and D. Gündüz, "Federated learning over wireless fading channels," *IEEE Trans. Wireless Commun.*, vol. 19, no. 5, pp. 3546–3557, May 2020.

[17] J. Sun, Y. Wang, X. Sun, N. Li, and G. Nie, "Time efficient joint optimization federated learning over wireless communication networks," *China Commun.*, vol. 19, no. 6, pp. 169–178, Jun. 2022.

[18] P. S. Bouzinis, P. D. Diamantoulakis, and G. K. Karagiannidis, "Wireless federated learning (WFL) for 6G networks—Part II: The compute-then-transmit NOMA paradigm," *IEEE Commun. Lett.*, vol. 26, no. 1, pp. 8–12, Jan. 2022.

[19] Y. Wu, Y. Song, T. Wang, L. Qian, and T. Q. S. Quek, "Non-orthogonal multiple access assisted federated learning via wireless power transfer: A cost-efficient approach," *IEEE Trans. Commun.*, vol. 70, no. 4, pp. 2853–2869, Apr. 2022.

[20] Z. Yang, M. Chen, W. Saad, C. S. Hong, and M. Shikh-Bahaei, "Energy efficient federated learning over wireless communication networks," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1935–1949, Mar. 2021.

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

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

[23] Q. Zeng, Y. Du, K. Huang, and K. K. Leung, "Energy-efficient resource management for federated edge learning with CPU-GPU heterogeneous computing," *IEEE Trans. Wireless Commun.*, vol. 20, no. 12, pp. 7947–7962, Dec. 2021.

[24] Q. Pham, M. Le, T. Huynh-The, Z. Han, and W. Hwang, "Energy-efficient federated learning over UAV-enabled wireless powered communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 5, pp. 4977–4990, May 2022.
[25] T. T. Vu, H. Quoc Ngo, T. L. Marzetta, and M. Matthaiou, "How does cell-free massive MIMO support multiple federated learning groups?" in *Proc. IEEE 22nd Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sep. 2021, pp. 401–405.
[26] N. Shende, O. Gurbuz, and E. Erkip, "Half-duplex or full-duplex relaying: A capacity analysis under self-interference," in *Proc. 47th Annu. Conf. Inf. Sci. Syst. (CISS)*, Mar. 2013, pp. 1–6.
[27] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.
[28] H. Q. Ngo and E. G. Larsson, "No downlink pilots are needed in TDD massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 16, no. 5, pp. 2921–2935, May 2017.
[29] E. Sharma, R. Budhiraja, K. Vasudevan, and L. Hanzo, "Full-duplex massive MIMO multi-pair two-way AF relaying: Energy efficiency optimization," *IEEE Trans. Commun.*, vol. 66, no. 8, pp. 3322–3340, Aug. 2018.
[30] S. Wang, Y. Liu, W. Zhang, and H. Zhang, "Achievable rates of full-duplex massive MIMO relay systems over Rician fading channels," *IEEE Trans. Veh. Technol.*, vol. 66, no. 11, pp. 9825–9837, Nov. 2017.
[31] C. Ma et al., "Distributed optimization with arbitrary local solvers," *Optim. Methods Softw.*, vol. 32, no. 4, pp. 813–848, Jul. 2017.
[32] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[33] L. D. Nguyen, H. D. Tuan, T. Q. Duong, H. V. Poor, and L. Hanzo, "Energy-efficient multi-cell massive MIMO subject to minimum user-rate constraints," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 914–928, Feb. 2021.
[34] Z. Sheng, H. D. Tuan, A. A. Nasir, T. Q. Duong, and H. V. Poor, "Power allocation for energy efficiency and secrecy of wireless interference networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 6, pp. 3737–3751, Jun. 2018.
[35] Z. Gülgün, E. Björnson, and E. G. Larsson, "Is massive MIMO robust against distributed jammers?" *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 457–469, Jan. 2021.
[36] T. T. Vu, D. T. Ngo, H. Q. Ngo, and T. Le-Ngoc, "Full-duplex cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019, pp. 1–6.
[37] H. Kim, J. Park, M. Bennis, and S.-L. Kim, "Blockchained on-device federated learning," *IEEE Commun. Lett.*, vol. 24, no. 6, pp. 1279–1283, Jun. 2020.

Muhammad Farooq (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the Department of Electrical Engineering, University of Engineering and Technology at Lahore, Pakistan, in 2014 and 2018, respectively, and the Ph.D. degree from the School of Electrical and Electronic Engineering, University College Dublin, Ireland, in February 2023.

He has worked as a Lecturer with the University of Engineering and Technology at Lahore, in 2018, and the University of Sargodha, Pakistan, from 2015 to 2018. He is currently a Post-Doctoral Researcher with the School of Computer Science, University College Dublin, Ireland. He has authored and coauthored conference papers in IEEE PIMRC in 2020, IWCMC in 2020, IEEE VTC in Spring 2021, EUSIPCO in 2021, and IEEE GLOBECOM in 2022, and journal articles in IEEE TRANSACTIONS ON COMMUNICATIONS and IEEE COMMUNICATIONS LETTERS. His research interests include wireless communication, signal processing, machine learning, optimization techniques, and intelligent transport systems. He also works as a Reviewer for IEEE TRANSACTIONS ON COMMUNICATIONS, IEEE SYSTEMS JOURNAL, IEEE COMMUNICATIONS LETTERS, IEEE WIRELESS COMMUNICATIONS LETTERS, and IEEE conferences.

Tung Thanh Vu (Member, IEEE) received the Ph.D. degree in wireless communications from The University of Newcastle, Australia, in 2021.

He is currently a Post-Doctoral Researcher with the Department of Electrical Engineering (ISY), Linköping University, Sweden. His research interests include optimization, information theories, and machine learning applications for 5G-and-beyond wireless networks, especially with massive MIMO, cell-free massive MIMO, federated learning, full-duplex communications, physical layer security, and low-earth orbit satellite communications. He received the Best Poster Award at AMSI Optimize Conference in 2018. He has also served as a member of the Technical Program Committee and the Symposium/Session Chair for a number of IEEE international conferences, such as GLOBECOM, ICCE, and ATC. He is serving as an Editor for *Physical Communication* (PHYCOM, Elsevier). He was an Exemplary Reviewer of IEEE WIRELESS COMMUNICATIONS LETTERS from 2020 to 2021 and IEEE TRANSACTIONS ON COMMUNICATIONS in 2021.

Hien Quoc Ngo (Senior Member, IEEE) is currently a Reader with Queen's University Belfast, U.K. He has coauthored many research articles in wireless communications and coauthored the textbook *Fundamentals of Massive MIMO* (Cambridge University Press, 2016). His research interests include massive MIMO systems, cell-free massive MIMO, physical layer security, and cooperative communications.

He received the IEEE ComSoc Stephen O. Rice Prize in 2015, the IEEE ComSoc Leonard G. Abraham Prize in 2017, and the Best Ph.D. Award from EURASIP in 2018. He also received the IEEE Sweden VT-COM-IT Joint Chapter Best Student Journal Paper Award in 2015. He was awarded the UKRI Future Leaders Fellowship in 2019. He serves as an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, the *Digital Signal Processing*, and the *Physical Communication* (Elsevier). He was a Guest Editor of *IET Communications* and IEEE ACCESS in 2017.

Le-Nam Tran (Senior Member, IEEE) received the B.S. degree in electrical engineering from the Ho Chi Minh City University of Technology, Ho Chi Minh City, Vietnam, in 2003, and the M.S. and Ph.D. degrees in radio engineering from Kyung Hee University, Seoul, South Korea, in 2006 and 2009, respectively.

He is currently an Assistant Professor with the School of Electrical and Electronic Engineering, University College Dublin, Ireland. Prior to this, he was a Lecturer with the Department of Electronic Engineering, Maynooth University, Ireland. From 2010 to 2014, he has held a post-doctoral positions with the Signal Processing Laboratory, ACCESS Linnaeus Centre, KTH Royal Institute of Technology, Stockholm, Sweden, and the Centre for Wireless Communications, University of Oulu, Finland. His research interests include the applications of optimization techniques for wireless communications design, energy-efficient communications, physical layer security, cloud radio access networks, cell-free massive MIMO, and reconfigurable intelligent surfaces. He was a recipient of the Career Development Award from Science Foundation Ireland in 2018. He has served on the technical program committees of several IEEE major conferences. He was the Symposium Co-Chair of Cognitive Computing and Networking Symposium of International Conference on Computing, Networking and Communication (ICNC 2016) and the Co-Chair of the Workshop on Scalable Massive MIMO Technologies for Beyond 5G at IEEE ICC in 2020. He was a co-recipient of the 2021 IEEE GLOBECOM Best Paper Award and the 2020 IEEE PIMRC 2020 Best Student Experimental Paper Award. He is an Associate Editor of IEEE COMMUNICATIONS LETTERS and *EURASIP Journal on Wireless Communications and Networking*.