Received 22 November 2021; accepted 9 December 2021. Date of publication 15 December 2021; date of current version 4 February 2022.

Digital Object Identifier 10.1109/OJCOMS.2021.3135153

Full-Duplex Cell-Free Massive MIMO Systems: Analysis and Decentralized Optimization

SOUMYADEEP DATTA [1,2], **DHEERAJ NAIDU AMUDALA** [2] **(Graduate Student Member, IEEE), EKANT SHARMA** [3] **(Member, IEEE), ROHIT BUDHIRAJA** [2], **AND SHIVENDRA S. PANWAR** [1] **(Fellow, IEEE)**

[1] Department of Electrical and Computer Engineering, New York University Tandon School of Engineering, Brooklyn, NY 11201, USA

[2] Department of Electrical Engineering, Indian Institute of Technology Kanpur, Kanpur 208016, India

[3] Department of Electronics and Communication Engineering, Indian Institute of Technology Roorkee, Roorkee 247667, India

CORRESPONDING AUTHOR: S. DATTA (e-mail: sdatta@nyu.edu)

This work was supported in part by the Visvesvaraya Ph.D. Scheme, Ministry of Electronics and Information Technology, Government of India under Grant MEITY-PHD-2721; in part by NYU Wireless; and in part by the NY State Center for Advanced Technology in Telecommunications (CATT).
Part of this work was published in the IEEE International Conference on Communications (ICC), 2021
[DOI: 10.1109/ICC42927.2021.9500917].

ABSTRACT Cell-free (CF) massive multiple-input-multiple-output (mMIMO) deployments are usually investigated with half-duplex nodes and high-capacity fronthaul links. To leverage the possible gains in throughput and energy efficiency (EE) of full-duplex (FD) communications, we consider a FD CF mMIMO system with *practical limited-capacity fronthaul links*. We derive closed-form spectral efficiency (SE) lower bounds for this system with maximum-ratio combining/maximum-ratio transmission processing and optimal uniform quantization. We then optimize the weighted sum EE (WSEE) via downlink and uplink power control by using a two-layered approach: the first layer formulates the optimization as a generalized convex program, while the second layer solves the optimization decentrally using the alternating direction method of multipliers. We analytically show that the proposed two-layered formulation yields a Karush-Kuhn-Tucker point of the original WSEE optimization. We numerically show the influence of weights on the individual EE of the users, which demonstrates the utility of the WSEE metric to incorporate heterogeneous EE requirements of users. We show that low fronthaul capacity reduces the number of users each AP can support, and the cell-free system, consequently, becomes user-centric.

INDEX TERMS Decentralized optimization, energy efficiency, full-duplex (FD), limited-capacity fronthaul.

I. INTRODUCTION

MASSIVE multiple-input-multiple-output (mMIMO) wireless systems employ a large number of antennas at the base stations (BSs), and achieve higher spectral efficiency (SE) and energy efficiency (EE) with relatively simple signal processing [1], [2]. Two distinct mMIMO variants are being investigated in the literature: i) co-located, wherein all antennas are located at one place [1]; and ii) distributed, wherein antennas are spread over a large area [2, and the references therein], [3]–[5]. While co-located mMIMO systems have a low fronthaul requirement, distributed mMIMO systems, at the cost of higher fronthaul infrastructure, have greater spatial diversity to exploit and consequently have greater immunity to shadow fading [2]–[4]. Cell-free (CF) mMIMO is one of the most promising distributed mMIMO variants in the current literature [2]–[5]. CF mMIMO envisions a communication region with no cell boundaries, and promises substantial gains in SE and fairness over small-cell deployments [3]–[5].

Full-duplex (FD) wireless systems have now been practically realized with advanced self-interference (SI) cancellation mechanisms [6]–[9]. Co-located FD massive MIMO systems have also been extensively investigated [10], [11, and the referencestherein]. FD CF mMIMO is a relatively recent area of interest [12]–[14], where access points (APs) simultaneously serve downlink and uplink user equipments (UEs) on the same spectral resource. Vu *et al.* in [12] considered a FD CF mMIMO system with maximum-ratio combining and showed that if SI at the APs is suppressed up to a certain limit, it has higher throughput than its half-duplex (HD) counterpart and FD co-located systems. Wang *et al.* in [13] evaluated the SE of a network-assisted FD CF mMIMO system using zero-forcing and regularized zero-forcing beamforming. Reference [14] proposed a heap-based algorithm for pilot assignment to overcome pilot contamination in FD CF mMIMO systems.

In CF mMIMO, APs are connected to a central processing unit (CPU) using fronthaul links. The existing FD CF mMIMO literature assumes high-capacity fronthaul links [12]–[14]. These links, however, have limited capacity, and the information needs to be consequently quantized and sent over them. The limited-capacity fronthaul has been considered only for HD CF mMIMO systems in [15]–[17]. Femenias and Riera-Palou in [16] studied a max-min uplink/downlink power allocation problem for HD CF mMIMO with limited-capacity fronthaul, while Masoumi and Emadi in [17] optimized the SE of a HD CF mMIMO uplink with limited-capacity fronthaul and hardware impairments. Bashar *et al.* in [15] derived the SE of HD CF mMIMO uplink with limited-capacity fronthaul. We consider quantized fronthaul for a FD CF mMIMO system to derive achievable SE expressions. To the best of our knowledge, the current work is first one to do so.

With tremendous increase in network traffic, the EE has become an important metric to design a modern wireless system. Global energy efficiency (GEE), defined as the ratio of the network SE and its total energy consumption, is being used to design CF mMIMO communication systems [18]–[21]. Ngo *et al.* in [18] optimized the GEE for the downlink of a HD CF mMIMO system. Bashar *et al.* in [19] optimized the uplink GEE of a HD CF mMIMO system with optimal uniform fronthaul quantization. Alonzo *et al.* in [20] optimized the GEE of CF and UE-centric HD mMIMO deployments in the mmWave regime. Nguyen *et al.* in [21] maximized a novel SE-GEE metric for the FD CF mMIMO system using a Dinkelbach-like algorithm.

A UE with limited energy availability will accord a much higher importance to its EE than an another UE with a sufficient energy supply. GEE is a network-centric metric and cannot accommodate such heterogeneous EE requirements [22]. The weighted sum energy efficiency (WSEE) metric, defined as the weighted sum of individual EEs [22], can prioritize EEs of individual UEs, by allocating them a higher weight [23], [24]. The WSEE is investigated in [23] for a general wireless network, and for a two-way FD relay in [24]. It is yet to be investigated for CF mMIMO HD and FD systems.

Decentralized designs, which accomplish a complex task by coordination and cooperation of a set of computing units, are being used to design mMIMO systems [25], [26]. This interest is driven by high computational complexity and high interconnection data rate requirements between radio frequency chains and baseband units in centralized mMIMO system designs [25]. Jeon *et al.* in [25] constructed decentralized equalizers by partitioning the BS antenna array. Reference [26] proposed a coordinate-descent-based decentralized algorithm for mMIMO uplink detection and downlink precoding. Reference [27] employed alternating direction method of multipliers (ADMM) to decentrally allocate edge-computing resource for vehicular networks. Such decentralized approaches have not yet been employed to optimize FD CF mMIMO systems. We next list our **main** contributions in this context:

1) *Contributions Regarding Closed Form SE Lower Bound:* We consider FD CF mMIMO communications with maximal ratio combining/maximal ratio transmission (MRC)/(MRT) processing and limited fronthaul with optimal uniform quantization. We note that for the FD CF mMIMO systems, *unlike their HD counterparts* [2]–[5], uplink and downlink transmissions interfere to cause *uplink downlink interference (UDI)* and *inter-/intra-AP residual interference (RI)*. Further, unlike existing FD CF mMIMO literature [12]–[14], [21], which consider perfect high-capacity fronthaul links, it is critical to model and analyze the UDI and inter-/intra-AP interferences and limited-capacity impairments while deriving lower bounds for both uplink and downlink UEs SE, which are valid for arbitrary number of antennas at each AP. We model the UDI on the downlink and the RI on the uplink, but unlike existing FD CF mMIMO literature [12]–[14], [21], we also consider the quantization distortion due to limited-capacity fronthaul links, as modelled in the total quantization distortion (TQD) terms. We also show the impact of quantization on the uplink RI terms themselves, where the distortion in the downlink and uplink signals get coupled. We derive achievable SE expressions for both uplink and downlink UEs, which are valid for arbitrary number of antennas at each AP.

2) *Contributions Regarding Centralized WSEE Optimization:* We use the derived SE expression to maximize the non-convex WSEE metric. While energy-efficient design of CF mMIMO systems have been studied in literature [18]–[20], most of them focus on the GEE metric, except [21]. The GEE, being a single ratio, can be expressed as a pseudo-concave (PC) function and can thus be maximized using Dinkelbach's algorithm [22]. Reference [21] is the only work so far which optimized the EE of FD CF mMIMO. It considered a novel SE-GEE objective, which also reduces to a PC function and is maximized using a Dinkelbach-like algorithm. *The WSEE, in contrast, is a sum of PC functions, and is not guaranteed to be a PC function* [22]. This makes the WSEE an extremely non-trivial objective to maximize [22]. Further, the algorithm in [21] requires knowledge of instantaneous small-scale channel fading coefficients. The WSEE metric optimized here, in contrast, requires large-scale channel coefficients, which remains constant for multiple coherence intervals [28].

3) *Contributions Regarding Decentralized Optimization:* We decentrally maximize WSEE using a two-layered

iterative approach which combines successive convex approximation (SCA) and ADMM. The first layer simplifies the non-convex WSEE maximization problem by using epigraph transformation, slack variables and series approximations. It then locally approximates the problem as a generalized convex program (GCP) which is solved iteratively using the SCA approach. The second layer decentrally optimizes the GCP by using the consensus ADMM approach, which decomposes the centralized version into multiple sub-problems, each of which is solved independently. The local solutions are combined to obtain the global solution. We note that the GCP for the FD system is not in the standard form which is required for applying ADMM, as it involves FD interference terms that couple power control coefficients from different UEs as well as from the uplink and downlink. We therefore create global and local versions of the power control coefficients separately for the downlink and uplink UEs, which decouple the FD interference terms. We consider separate sub-problems for the downlink and uplink UEs with a separate set of constraints for each. These constraints, rewritten using the local variables, define feasible sets for the sub-problems of the downlink and uplink UEs, respectively. We introduce separate Lagrangian parameters for the downlink and uplink UEs, and separate penalty parameters for the downlink and uplink power control variables. This enables us to properly define the augmented Lagrangian and decouple the respective sub-problems at the D-servers which calculate the local solutions, and then eventually coordinate them into the globally optimal solution at the C-server. *The FD system required that we introduce these modifications to the standard ADMM approach and to the best of our knowledge, has not been attempted so far in mMIMO literature.*

4) *Contributions Regarding the AP Selection Algorithm:* We show that there is a fundamental limit to the number of UEs a FD AP can serve with a limited fronthaul capacity. We propose a *proportionately-fair* rule capping the maximum number of uplink and downlink UEs served by each AP. We use this rule to propose a fair AP selection algorithm which efficiently chooses the best subset of APs to serve each uplink and downlink UE. The proposed approach ensures user-centric architecture for our system. The proposed algorithm, which has a trivial complexity, is shown to perform close to the optimal one proposed in [29].

5) *Contributions Regarding the Convergence of the Distributed Optimization Algorithm:* We not only analytically prove its convergence but also numerically show that it i) achieves the same WSEE as the centralized approach; and ii) is responsive to changing weights which can be set to prioritize UEs' EE requirements.

II. SYSTEM MODEL

We consider, as shown in Fig. 1, a FD CF mMIMO system where M FD APs serve $K = (K_u + K_d)$ single-antenna HD UEs on the same spectral resource, with K_u and K_d being the number of uplink and downlink UEs, respectively. Each AP has N_t transmit and N_r receive antennas, and is connected to the CPU using a limited-capacity fronthaul link which carries quantized uplink/downlink information to/from the CPU. We see from Fig. 1 that due to FD model

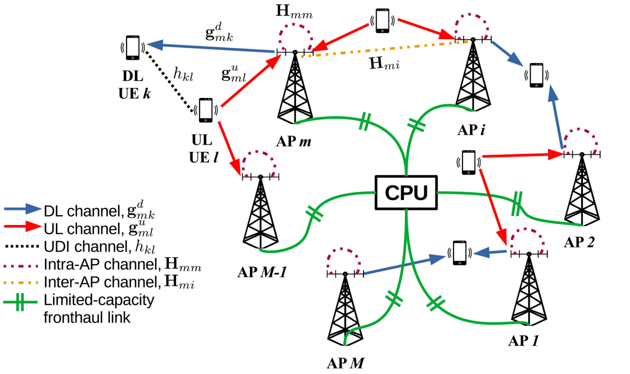

$\mathbf{H}_{mm}$
$\mathbf{g}^d_{mk}$
DL UE *k*
h_{kl}
$\mathbf{g}^u_{ml}$
UL UE *l*
AP *m*
$\mathbf{H}_{mi}$
AP *i*
CPU
AP *2*
AP *M-1*
AP *M*
AP *1*
DL channel, $\mathbf{g}^d_{mk}$
UL channel, $\mathbf{g}^u_{ml}$
UDI channel, h_{kl}
Intra-AP channel, $\mathbf{H}_{mm}$
Inter-AP channel, $\mathbf{H}_{mi}$
Limited-capacity fronthaul link

FIGURE 1. **System model for FD CF mMIMO communications.**

- uplink receive signal of each AP is interfered by its own downlink transmit signal and that of other APs. These intra- and inter-AP interferences are shown using purple and brown dashed lines, respectively.
- downlink UEs receive transmit signals from uplink UEs, causing uplink downlink interference (UDI) (shown as black dotted lines between uplink and downlink UEs). Additionally, the UEs experience multi-UE interference (MUI) as the APs serve them on the same spectral resource.

We next explain various channels, their estimation and data transmission. We assume a coherence interval of duration T_c (in s) with τ_c samples, which is divided into: a) channel estimation phase of τ_t samples, and b) downlink and uplink data transmission of (τ_c - τ_t) samples.

A. CHANNEL DESCRIPTION

The channel of the kth downlink UE to the transmit antennas of the mth AP is $\boldsymbol{g}^d_{mk} \in \mathbb{C}^{N_t \times 1}$, while the channel from the lth uplink UE to the receive antennas of the mth AP is $\boldsymbol{g}^u_{ml} \in \mathbb{C}^{N_r \times 1}$.[1] We model these channels as $\boldsymbol{g}^d_{mk} = (\beta^d_{mk})^{1/2}\tilde{\boldsymbol{g}}^d_{mk}$ and $\boldsymbol{g}^u_{ml} = (\beta^u_{ml})^{1/2}\tilde{\boldsymbol{g}}^u_{ml}$. Here β^d_{mk} and $\beta^u_{ml} \in \mathbb{R}$ are corresponding large scale fading coefficients, which are same for all antennas at the mth AP [3], [12]. The vectors $\tilde{\boldsymbol{g}}^d_{mk}$ and $\tilde{\boldsymbol{g}}^u_{ml}$ denote small scale fading with independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ entries. The UDI channel between the kth downlink UE and lth uplink UE is modeled as $h_{kl} = (\tilde{\beta}_{kl})^{1/2}\tilde{h}_{kl}$ [12], [13], where $\tilde{\beta}_{kl}$ is the large scale fading coefficient and $\tilde{h}_{kl} \sim \mathcal{CN}(0, 1)$ is the small scale fading. The inter- and intra-AP channels from the transmit antennas of the ith AP to the receive antennas of the mth AP are denoted as $\boldsymbol{H}_{mi} \in \mathbb{C}^{N_r \times N_t}$ for $i = 1$ to M.

B. UPLINK CHANNEL ESTIMATION

Recall that the channel estimation phase consists of τ_t samples. We divide them as $\tau_t = \tau^d_t + \tau^u_t$, where τ^d_t and τ^u_t

[1] We, henceforth, consider $k = 1$ to K_d, $l = 1$ to K_u and $m = 1$ to M, to avoid repetition, unless mentioned otherwise.

are samples used as pilots for the downlink and uplink UEs, respectively. All the downlink (resp. uplink) UEs simultaneously transmit τ_t^d (resp. τ_t^u)-length uplink pilots to the APs, which they use to estimate the respective channels. In this phase, both transmit and receive antenna arrays of each AP, similar to [12], operate in receive mode. The kth downlink UE (resp. lth uplink UE) transmits pilot signals $\sqrt{\tau_t^d}\boldsymbol{\varphi}_k^d \in \mathbb{C}^{\tau_t^d\times 1}$ (resp. $\sqrt{\tau_t^u}\boldsymbol{\varphi}_l^u \in \mathbb{C}^{\tau_t^u\times 1}$). We assume, similar to [12], [18], that the pilots i) have unit norm, i.e., $\left\|\boldsymbol{\varphi}_l^u\right\| = \left\|\boldsymbol{\varphi}_k^d\right\| = 1$; and ii) are intra-set orthonormal, i.e., $(\boldsymbol{\varphi}_l^u)^H\boldsymbol{\varphi}_{l'}^u = 0\,\forall l \neq l'$ and $(\boldsymbol{\varphi}_k^d)^H\boldsymbol{\varphi}_{k'}^d = 0\,\forall k \neq k'$. Therefore, we need $\tau_t^d \geq K_d$ and $\tau_t^u \geq K_u$ [12], [18].

The pilots received by transmit and receive antennas of the mth AP are given respectively as

$$\boldsymbol{Y}_m^{tx} = \sqrt{\tau_t^d\rho_t}\sum_{k=1}^{K_d}\boldsymbol{g}_{mk}^d\left(\boldsymbol{\varphi}_k^d\right)^H + \boldsymbol{W}_m^{tx},$$

$$\boldsymbol{Y}_m^{rx} = \sqrt{\tau_t^u\rho_t}\sum_{l=1}^{K_u}\boldsymbol{g}_{ml}^u\left(\boldsymbol{\varphi}_l^u\right)^H + \boldsymbol{W}_m^{rx}.$$

Here ρ_t is the normalized pilot transmit signal-to-noise-ratio (SNR). The matrices $\boldsymbol{W}_m^{tx} \in \mathbb{C}^{N_t\times\tau_t^d}$ and $\boldsymbol{W}_m^{rx} \in \mathbb{C}^{N_r\times\tau_t^u}$ denote additive noise with $\mathcal{CN}(0,1)$ entries. Each AP independently estimates its channels with the uplink and downlink UEs to avoid channel state information (CSI) exchange overhead [12], [21]. To estimate the channels $\boldsymbol{g}_{mk}^d$ and $\boldsymbol{g}_{ml}^u$, the mth AP projects the received signal onto the pilot signals $\boldsymbol{\varphi}_k^d$ and $\boldsymbol{\varphi}_l^u$ respectively, as

$$\hat{\boldsymbol{y}}_{mk}^{tx} = \boldsymbol{Y}_m^{tx}\boldsymbol{\varphi}_k^d = \sqrt{\tau_t^d\rho_t}\boldsymbol{g}_{mk}^d + \boldsymbol{W}_m^{tx}\boldsymbol{\varphi}_k^d$$

$$\hat{\boldsymbol{y}}_{ml}^{rx} = \boldsymbol{Y}_m^{rx}\boldsymbol{\varphi}_l^u = \sqrt{\tau_t^u\rho_t}\boldsymbol{g}_{ml}^u + \boldsymbol{W}_m^{rx}\boldsymbol{\varphi}_l^u.$$

These projections are used to compute the corresponding linear minimum-mean-squared-error (MMSE) channel estimates [12] as

$$\hat{\boldsymbol{g}}_{mk}^d = \mathbb{E}\left\{\boldsymbol{g}_{mk}^d\left(\hat{\boldsymbol{y}}_{mk}^{tx}\right)^H\right\}\left(\mathbb{E}\left\{\hat{\boldsymbol{y}}_{mk}^{tx}\left(\hat{\boldsymbol{y}}_{mk}^{tx}\right)^H\right\}\right)^{-1}\hat{\boldsymbol{y}}_{mk}^{tx} = c_{mk}^d\hat{\boldsymbol{y}}_{mk}^{tx},$$

$$\hat{\boldsymbol{g}}_{ml}^u = \mathbb{E}\left\{\boldsymbol{g}_{ml}^u\left(\hat{\boldsymbol{y}}_{ml}^{rx}\right)^H\right\}\left(\mathbb{E}\left\{\hat{\boldsymbol{y}}_{ml}^{rx}\left(\hat{\boldsymbol{y}}_{ml}^{rx}\right)^H\right\}\right)^{-1}\hat{\boldsymbol{y}}_{ml}^{rx} = c_{ml}^u\hat{\boldsymbol{y}}_{ml}^{rx},$$

where $c_{mk}^d = \frac{\sqrt{\tau_t^d\rho_t}\beta_{mk}^d}{\tau_t^d\rho_t\beta_{mk}^d+1}$ and $c_{ml}^u = \frac{\sqrt{\tau_t^u\rho_t}\beta_{ml}^u}{\tau_t^u\rho_t\beta_{ml}^u+1}$. The estimation error vectors are defined as $\boldsymbol{e}_{ml}^u \triangleq \boldsymbol{g}_{ml}^u - \hat{\boldsymbol{g}}_{ml}^u$ and $\boldsymbol{e}_{mk}^d \triangleq \boldsymbol{g}_{mk}^d - \hat{\boldsymbol{g}}_{mk}^d$. With MMSE channel estimation, $\hat{\boldsymbol{g}}_{mk}^d, \boldsymbol{e}_{mk}^d$ and $\hat{\boldsymbol{g}}_{ml}^u, \boldsymbol{e}_{ml}^u$ are mutually independent and their individual terms are i.i.d. with pdf $\mathcal{CN}(0,\gamma_{mk}^d), \mathcal{CN}(0,\beta_{mk}^d - \gamma_{mk}^d), \mathcal{CN}(0,\gamma_{ml}^u), \mathcal{CN}(0,\beta_{ml}^u - \gamma_{ml}^u)$ respectively, with $\gamma_{mk}^d = \frac{\tau_t^d\rho_t(\beta_{mk}^d)^2}{\tau_t^d\rho_t\beta_{mk}^d+1}$ and $\gamma_{ml}^u = \frac{\tau_t^u\rho_t(\beta_{ml}^u)^2}{\tau_t^u\rho_t\beta_{ml}^u+1}$ [12], [18].

After channel estimation, data transmission starts simultaneously on downlink and uplink.

C. TRANSMISSION MODEL

An objective of this work is to derive a SE lower bound for FD CF mMIMO systems, where the M APs serve K_u uplink UEs and K_d downlink UEs simultaneously on the same spectral resource. We note that for the FD CF mMIMO systems, unlike the HD CF mMIMO systems [3], [15], [16], uplink and downlink transmissions interfere to cause UDI and inter-/intra-AP interferences. Further, unlike existing FD CF mMIMO literature [12], [13], [21], we consider a limited-capacity fronthaul. It is critical to model and analyze the UDI and inter-/intra-AP interferences and limited-capacity impairments while deriving the lower bound.

1) DOWNLINK DATA TRANSMISSION

The CPU chooses a message symbol s_k^d for the kth downlink UE, which is distributed as $\mathcal{CN}(0,1)$. It intends to send this symbol to the mth AP via the limited-capacity fronthaul link. Before doing that, it multiplies s_k^d with a power-control coefficient η_{mk}, and then quantizes the resulting signal. The mth AP, due to its limited fronthaul capacity, is allowed to serve only a subset $\kappa_{dm} \subset \{1,\ldots,K_d\}$ of downlink users, an aspect which is discussed later in Section II-D. The CPU consequently sends downlink symbols for UEs in the set κ_{dm} to the mth AP, which uses MMSE channel estimates to perform MRT precoding. The transmit signal of the mth AP is therefore given as follows

$$\begin{aligned}\boldsymbol{x}_m^d &= \sqrt{\rho_d}\sum_{k\in\kappa_{dm}}\left(\hat{\boldsymbol{g}}_{mk}^d\right)^*\mathcal{Q}\left(\sqrt{\eta_{mk}}s_k^d\right)\\ &= \sqrt{\rho_d}\sum_{k\in\kappa_{dm}}\left(\hat{\boldsymbol{g}}_{mk}^d\right)^*\left(\tilde{a}\sqrt{\eta_{mk}}s_k^d + \varsigma_{mk}^d\right).\end{aligned} \tag{1}$$

Here ρ_d is the normalized maximum transmit SNR at each AP. The function $\mathcal{Q}(\cdot)$ denotes the quantization operation, which is modeled as a multiplicative attenuation $\tilde{a}$, and an additive distortion ς_{mk}^d, for the kth downlink UE in the fronthaul link between the CPU and the mth AP [15], [19]. We have, from Appendix A, $\mathbb{E}\{(\varsigma_{mk}^d)^2\} = (\tilde{b} - \tilde{a}^2)\mathbb{E}\{|\sqrt{\eta_{mk}}s_k^d|^2\} = (\tilde{b} - \tilde{a}^2)\eta_{mk}$, where the scalar constants $\tilde{a}$ and $\tilde{b}$ depend on the number of fronthaul quantization bits.

The mth AP must satisfy the average transmit SNR constraint, i.e., $\mathbb{E}\{\|\boldsymbol{x}_m^d\|^2\} \leq \rho_d$. Using the expression of $\boldsymbol{x}_m^d$ from (1), and the above expression of quantization error variance, $\mathbb{E}\{(\varsigma_{mk}^d)^2\}$, the constraint can be simplified as follows

$$\rho_d\tilde{b}\sum_{k\in\kappa_{dm}}\eta_{mk}\mathbb{E}\left\{\|\hat{\boldsymbol{g}}_{mk}^d\|^2\right\} \leq \rho_d \Rightarrow \tilde{b}\sum_{k\in\kappa_{dm}}\gamma_{mk}^d\eta_{mk} \leq \frac{1}{N_t}. \tag{2}$$

The kth downlink UE receives its desired message signal from a subset of all APs, denoted as $\mathcal{M}_k^d \subset \{1,\ldots,M\}$, along with various interference and distortion components, as in (5) (shown at the top of the next page). The mth AP serves the kth downlink UE iff $k \in \kappa_{dm} \Leftrightarrow m \in \mathcal{M}_k^d$. Here x_l^u is the transmit signal of the lth uplink UE, which is modelled next.

2) UPLINK DATA TRANSMISSION

The K_u uplink UEs also simultaneously transmit to all M APs on the same spectral resource as that of the K_d downlink UEs. The lth uplink UE transmits its signal $x_l^u = \sqrt{\rho_u\theta_l}s_l^u$ with

s_l^u being its message symbol with pdf $\mathcal{CN}(0,1)$, ρ_u being the maximum uplink transmit SNR and θ_l being the power control coefficient. To satisfy the average SNR constraint, $\mathbb{E}\{|x_l^u|^2\} \leq \rho_u$, the lth uplink UE satisfies the constraint

$$0 \leq \theta_l \leq 1. \tag{3}$$

The FD APs not only receive the uplink UE signals but also their own downlink transmit signals and that of the other APs, referred to as intra-AP and inter-AP interference, respectively. Using (1), the received uplink signal at the mth AP is

$$\boldsymbol{y}_m^u = \sum_{l=1}^{K_u} \boldsymbol{g}_{ml}^u x_l^u + \sum_{i=1}^{M} \boldsymbol{H}_{mi}\boldsymbol{x}_i^d + \boldsymbol{w}_m^u = \sqrt{\rho_u}\sum_{l=1}^{K_u} \boldsymbol{g}_{ml}^u \sqrt{\theta_l} s_l^u \\ + \sqrt{\rho_d}\sum_{i=1}^{M}\sum_{k\in\kappa_{di}} \boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\left(\tilde{a}\sqrt{\eta_{ik}}s_k^d + \varsigma_{ik}^d\right) + \boldsymbol{w}_m^u. \tag{4}$$

Here $\boldsymbol{w}_m^u \in \mathbb{C}^{N_r\times 1}$ is the additive receiver noise at the mth AP with i.i.d. entries $\sim \mathcal{CN}(0,1)$.

The intra and inter-AP interference channels vary extremely slowly and thus can be estimated with very low pilot overhead [13]. The receive antenna array of each AP, with estimated channel, can only partially mitigate the intra- and inter-AP interference [12], [13]. The residual intra-/inter-AP interference (RI) channel $\boldsymbol{H}_{mi} \in \mathbb{C}^{N_r\times N_t}$ is modeled as Rayleigh-faded with i.i.d. entries and pdf $\mathcal{CN}(0, \gamma_{\text{RI},mi})$ [6], [12], [13], [24]. Here $\gamma_{\text{RI},mi} \triangleq \beta_{\text{RI},mi}\gamma_{\text{RI}}$, with $\beta_{\text{RI},mi}$ being the large scale fading coefficient from the ith AP to the mth AP, and γ_{RI} being the RI power after its suppression.

The mth AP receives the signals from all the uplink UEs, and performs MRC for the lth uplink UE with $(\hat{\boldsymbol{g}}_{ml}^u)^H$. Due to its limited fronthaul: i) AP quantizes the combined signal before sending it to CPU; ii) as discussed in detail later in Section II-D, the CPU receives contributions for the lth uplink UE only from the subset of APs serving it, denoted as $\mathcal{M}_l^u \subset \{1,\ldots,M\}$. Using (4), the signal received by the CPU for the lth uplink UE is expressed as in (6) (shown at the top of the next page).

We denote the subset of uplink UEs served by the mth AP as $\kappa_{um} \subset \{1,\ldots,K_u\}$. The mth AP serves the lth uplink UE iff $l \in \kappa_{um} \Leftrightarrow m \in \mathcal{M}_l^u$. The quantization operation $\mathcal{Q}(\cdot)$ is mathematically modeled using constant attenuation $\tilde{a}$, and additive distortion ς_{ml}^u which, as shown in Appendix A, has power $\mathbb{E}\{(\varsigma_{ml}^u)^2\} = (\tilde{b}-\tilde{a}^2)\mathbb{E}\{|(\boldsymbol{g}_{ml}^u)^H\boldsymbol{y}_m|^2\}$.

D. USER-CENTRIC BEHAVIOR THROUGH LIMITED FRONTHAUL

Initial CF mMIMO literature considered system models where all APs can serve all UEs [3]–[5]. However, for geographically large areas, each UE can only have practically feasible channels with a subset of APs in its vicinity. Therefore, recent CF mMIMO literature has increasingly focused on user-centric CF mMIMO system design [2, and the references therein]. In the subsequent discussion, we show that a user-centric CF deployment, as desired by us, is a natural outcome of the design choice to impose fronthaul capacity constraints on the CF mMIMO system model, as shown in Fig. 1.

The fronthaul between the mth AP and the CPU uses ν_m bits to quantize the real and imaginary parts of transmit signal of the mth downlink UE and the uplink receive signal after MRC, i.e., $\sqrt{\eta_{mk}}s_k^d$, and $(\hat{\boldsymbol{g}}_{ml}^u)^H\boldsymbol{y}_m^u$, respectively. Due to the limited-capacity fronthaul, the mth AP serves only $K_{um}(\triangleq |\kappa_{um}|)$ and $K_{dm}(\triangleq |\kappa_{dm}|)$ UEs on the uplink and downlink, respectively [15], [19]. For each UE, we recall that there are $(\tau_c - \tau_t)$ data samples in each coherence interval of duration T_c. The fronthaul data rate between the mth AP and the CPU is

$$R_{\text{fh},m} = \frac{2\nu_m(K_{dm}+K_{um})(\tau_c-\tau_t)}{T_c}. \tag{7}$$

The fronthaul link between the mth AP and the CPU has capacity $C_{\text{fh},m}$ which implies that

$$R_{\text{fh},m} \leq C_{\text{fh},m} \Rightarrow \nu_m\cdot(K_{um}+K_{dm}) \leq \frac{C_{\text{fh},m}T_c}{2(\tau_c-\tau_t)}. \tag{8}$$

We propose the following lemma where we consider a *proportionally fair* approach to calculate K_{dm} and K_{um} *in proportion to* the total number of downlink and uplink UEs, respectively. We use $\varepsilon \triangleq \{d,u\}$ to denote downlink and uplink, respectively, and define the total number of UEs, $K \triangleq K_u + K_d$.

Lemma 1: The maximum number of uplink and downlink UEs served by the mth AP when connected via a limited optical fronthaul to the CPU with capacity $C_{\text{fh},m}$ are given as

$$\bar{K}_{\epsilon m} = \left\lfloor \frac{K_\epsilon}{K}\frac{C_{\text{fh},m}T_c}{4(\tau_c-\tau_t)\nu_m}\right\rfloor. \tag{9}$$

Proof: Let $\bar{K}_{um}$ and $\bar{K}_{dm}$ denote the maximum number of uplink and downlink UEs served by the mth AP. We consider $\bar{K}_{um} \propto K_u$ and $\bar{K}_{dm} \propto K_d$ for proportional fairness on the uplink and downlink. Using (8), we get,

$$\bar{K}_{\epsilon m} \leq \frac{K_\epsilon}{K}\frac{C_{\text{fh},m}T_c}{2(\tau_c-\tau_t)\nu_m}.$$

The lemma follows from definition of floor function $\lfloor\cdot\rfloor$. ■

Using the maximum limits obtained in (9), we assign $K_{um} = \min\{K_u, \bar{K}_{um}\}$ and $K_{dm} = \min\{K_d, \bar{K}_{dm}\}$. We see that the constraint imposed in (8) is similar to a UE-centric (UC) CF mMIMO system, wherein each UE is served by a subset of the APs [2]. We now define the procedure for AP selection to obtain the best subset of APs to serve each uplink and downlink UE, while satisfying (8). For this, we extend the procedure in [15] for a FD system as follows:

- The mth AP sorts the uplink and downlink UEs connected to it in descending order based on their channel gains (β_{ml}^u and β_{mk}^d, respectively) and chooses K_{um} uplink UEs and K_{dm} downlink UEs, with the largest channel gains, to populate the sets κ_{um} and κ_{dm}, respectively.

Algorithm 1: Fair AP Selection for Disconnected Uplink and Downlink UEs

```
for $k \leftarrow 1$ to $K_d$ do
    if $\mathcal{M}_k^d = \phi$ then
        Sort the APs in descending order of channel gains, $\beta_{mk}^d$, and find the AP $n$ with the largest channel gain.
        For this $n$th AP, sort downlink UEs in $\kappa_{dn}$ in descending order of channel gains and find the $q$th downlink UE with minimum channel gain and at least one more connected AP.
        Remove the $q$th downlink UE from the set $\kappa_{dn}$ and add the $k$th downlink UE to it.
Repeat the same procedure for all the uplink UEs $l = 1$ to $K_u$.
```

- For the lth uplink UE and the kth downlink UE, we populate the sets $\mathcal{M}_l^u$ and $\mathcal{M}_k^d$, respectively, using the axioms $l \in \kappa_{um} \Leftrightarrow m \in \mathcal{M}_l^u$ and $k \in \kappa_{dm} \Leftrightarrow m \in \mathcal{M}_k^d$.
- If an uplink or downlink UE is found with no serving AP, we use the procedure in Algorithm 1 to assign it the AP with the best channel conditions, while satisfying (8).

Clearly, Lemma II-D ensures that each AP can only serve a limited number of UEs which do not violate the fronthaul capacity constraints. This makes the system effectively a user-centric system. Algorithm 1 ensures that, under limited fronthaul constraints, the strongest AP-UE connections are retained and the UE-centric cell-free system delivers good performance.

E. SELF-INTERFERENCE MITIGATION METHODS

To ensure that our proposed FD CF mMIMO system has substantial performance improvement over an equivalent HD CF mMIMO system, we need effective techniques to cancel the self-interference (SI) caused due to inter-AP transmissions. We show in Eq. (5)-(6), shown at the bottom of the page that this SI cancellation results in a residual interference (RI) due to the multiplication of a suppression factor, γ_{RI}. We now discuss SI cancellation techniques from the existing literature, which makes the SI suppression easier, by not requiring its instantaneous channel knowledge.

- *Passive cancellation:* Reference [7], [30] suggests that a careful utilization of the passive self-interference suppression mechanisms (directional isolation, absorptive shielding, and cross polarization) can significantly suppress the SI. Reference [7] also showed that by additionally assuming statistical SI channel knowledge and by using antennas arrays of sources/destinations, the passive cancellation techniques can further suppress the SI.
- *Large antenna array:* Reference [31] argued that with large N, channel vectors of the desired signal and the SI become nearly orthogonal. The beamforming techniques, e.g., MRC/MRT inherently project the desired signal to the orthogonal complement space of the SI, which significantly reduces the SI.
- *Lower transmit power:* Reference [31] also demonstrated that an alternative way to reduce interference could be to reduce transmit power, since the SI depends strongly on the transmit power. A cell free massive MIMO system, due to large number of transmit antennas, uses radically less transmit power/antenna than conventional MIMO systems, which significantly reduces the SI.
- We therefore, similar to existing massive MIMO FD literature [7], [31], [32], assume that the SI can be significantly mitigated by utilizing the above mentioned SI cancellation techniques, and without requiring the knowledge of SI channel. However, if required, the residual SI can be further reduced by employing active (time-domain and spatial suppression) techniques developed in [33], which require SI channel knowledge.

$$
\begin{aligned}
r_k^d = \sum_{m=1}^{M} \left(\boldsymbol{g}_{mk}^d\right)^T \boldsymbol{x}_m^d + \sum_{l=1}^{K_u} h_{kl} x_l^u + w_k^d = {} & \underbrace{\tilde{a}\sqrt{\rho_d} \sum_{m \in \mathcal{M}_k^d} \sqrt{\eta_{mk}} \left(\boldsymbol{g}_{mk}^d\right)^T \left(\hat{\boldsymbol{g}}_{mk}^d\right)^* s_k^d}_{\text{message signal}} + \underbrace{\tilde{a}\sqrt{\rho_d} \sum_{m=1}^{M} \sum_{q \in \kappa_{dm} \backslash k} \sqrt{\eta_{mq}} \left(\boldsymbol{g}_{mk}^d\right)^T \left(\hat{\boldsymbol{g}}_{mq}^d\right)^* s_q^d}_{\text{multi-UE interference, MUI}_k^d} \\
& + \underbrace{\sum_{l=1}^{K_u} h_{kl} x_l^u}_{\text{uplink downlink interference, UDI}_k^d} + \underbrace{\sqrt{\rho_d} \sum_{m=1}^{M} \sum_{q \in \kappa_{dm}} \left(\boldsymbol{g}_{mk}^d\right)^T \left(\hat{\boldsymbol{g}}_{mq}^d\right)^* \varsigma_{mq}^d}_{\text{total quantization distortion, TQD}_k^d} + \underbrace{w_k^d}_{\text{AWGN at receiver}}
\end{aligned}
\tag{5}
$$

$$
\begin{aligned}
r_l^u = \sum_{m \in \mathcal{M}_l^u} \mathcal{Q}\left(\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{y}_m^u\right) = {} & \underbrace{\tilde{a} \sum_{m \in \mathcal{M}_l^u} \sqrt{\rho_u}\sqrt{\theta_l} \left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{g}_{ml}^u s_l^u}_{\text{message signal}} + \underbrace{\tilde{a} \sum_{m \in \mathcal{M}_l^u} \sum_{q=1, q \neq l}^{K_u} \sqrt{\rho_u}\sqrt{\theta_q} \left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{g}_{mq}^u s_q^u}_{\text{multi-UE interference, MUI}_l^u} \\
& + \underbrace{\tilde{a} \sum_{m \in \mathcal{M}_l^u} \sum_{i=1}^{M} \sqrt{\rho_d} \sum_{k \in \kappa_{di}} \left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{H}_{mi} \left(\hat{\boldsymbol{g}}_{ik}^d\right)^* \left(\tilde{a}\sqrt{\eta_{ik}} s_k^d + \varsigma_{ik}^d\right)}_{\text{residual interference (intra-AP and inter-AP), RI}_l^u} + \underbrace{\tilde{a} \sum_{m \in \mathcal{M}_l^u} \left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{w}_m^u}_{\text{AWGN at APs, N}_l^u} + \underbrace{\sum_{m \in \mathcal{M}_l^u} \varsigma_{ml}^u}_{\text{total quantization distortion, TQD}_l^u}
\end{aligned}
\tag{6}
$$

- *Active cancellation:* The authors in [33] present an algorithm for SI channel estimation at the relay, which is equipped with large number of antennas. It also noted that the APs, which are infrastructure devices, are in a stationary environment. The SI channel changes much more slowly than the channel from users to the APs. It is therefore reasonable to assume that i) the SI channel remains constant for multiple consecutive blocks; and ii) inter-AP pilot overhead is affordable because of the sufficiently longer coherence time of the residual SI channels. Similar to [33], one can estimate the SI channel by utilizing its slowly-varying nature using a cost-efficient expectation-maximization algorithm with reduced complexity.

III. ACHIEVABLE SPECTRAL EFFICIENCY

We now derive the ergodic SE for the kth downlink UE and the lth uplink UE, denoted respectively as $\bar{S}_k^d$ and $\bar{S}_l^u$. The AP employs MRC/MRT in the uplink/downlink and optimal uniform fronthaul quantization. We use $\varepsilon \triangleq \{d, u\}$ to denote downlink and uplink, respectively; $\phi \triangleq \{k, l\}$ to denote kth downlink UE and lth uplink UE, respectively; and $\upsilon_{m\phi}^{\varepsilon} \triangleq \{\eta_{mk}$ for $\phi = k, \theta_l$ for $\phi = l\}$. The ergodic SE expressions are calculated using (5) and (6), as

$$\begin{aligned}
&\bar{S}_{\phi}^{\varepsilon} = \left(\frac{\tau_c - \tau_t}{\tau_c}\right)\mathbb{E}\left\{\log_2\left(1 + \frac{P_{\phi}^{\varepsilon}}{I_{\phi}^{\varepsilon} + \left(\sigma_{\phi,0}^{\varepsilon}\right)^2}\right)\right\}, \text{ where}\\
&P_{\phi}^{\varepsilon} = \left|\tilde{a}\sum_{m\in\mathcal{M}_{\phi}^{\varepsilon}} \sqrt{\rho_{\varepsilon}}\sqrt{\upsilon_{m\phi}^{\varepsilon}}\left(\hat{\boldsymbol{g}}_{m\phi}^{\varepsilon}\right)^H \boldsymbol{g}_{m\phi}^{\varepsilon} s_{\phi}^{\varepsilon}\right|^2,\\
&\left(\sigma_{k,0}^{d}\right)^2 = \left|w_k^d\right|^2, \left(\sigma_{l,0}^{u}\right)^2 = \left|\tilde{a}\sum_{m\in\mathcal{M}_l^u}\left(\hat{\boldsymbol{g}}_{ml}^{u}\right)^H \boldsymbol{w}_m^u\right|^2,\\
&I_k^d = \left|\sum_{l=1}^{K_u} h_{kl}\sqrt{\rho_u\theta_l}s_l^u\right|^2\\
&\quad + \left|\tilde{a}\sqrt{\rho_d}\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}\backslash k}\sqrt{\eta_{mq}}\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mq}^d\right)^* s_q^d\right|^2\\
&\quad + \left|\sqrt{\rho_d}\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}}\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mq}^d\right)^* \varsigma_{mq}^d\right|^2,\\
&I_l^u = \left|\tilde{a}\sum_{m\in\mathcal{M}_l^u}\sum_{q=1,q\neq l}^{K_u}\sqrt{\rho_u}\sqrt{\theta_q}\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{g}_{mq}^u s_q^u\right|^2 + \left|\sum_{m\in\mathcal{M}_l^u}\varsigma_{ml}^u\right|^2\\
&\quad + \left|\tilde{a}\sum_{m\in\mathcal{M}_l^u}\sum_{i=1}^{M}\sqrt{\rho_d}\sum_{k\in\kappa_{di}}\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\left(\tilde{a}\sqrt{\eta_{ik}}s_k^d + \varsigma_{ik}^d\right)\right|^2,
\end{aligned}\tag{10}$$

are signal, noise and interference powers respectively, for the kth downlink and lth uplink UEs. The expectation outside logarithm in the SE expressions in (10) is mathematically intractable, and it is difficult to simplify them further [3], [12], [15]. We, similar to [3], employ use-and-then-forget (UatF) technique to derive SE lower bounds. To use UatF, we rewrite the received signal at the kth downlink UE in (5), and at the CPU for the lth uplink UE in (6) as

$$r_{\phi}^{\varepsilon} = \underbrace{\tilde{a}\sum_{m\in\mathcal{M}_{\phi}^{\varepsilon}}\sqrt{\rho_{\varepsilon}}\sqrt{\upsilon_{m\phi}^{\varepsilon}}\mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{m\phi}^{\varepsilon}\right)^H\boldsymbol{g}_{m\phi}^{\varepsilon}\right\}s_{\phi}^{\varepsilon}}_{\text{desired signal, } \mathrm{DS}_{\phi}^{\varepsilon}} + n_{\phi}^{\varepsilon}, \tag{11}$$

where the effective additive noise terms n_{ϕ}^{ε} are expressed in (12)-(13) (shown at the bottom of the page). The term $\mathrm{DS}_{\phi}^{\varepsilon}$ in (11) denotes the desired signal received over the channel mean, and the term $\mathrm{BU}_{\phi}^{\varepsilon}$ in (12)-(13) denotes beamforming uncertainty, i.e., the signal received over deviation of channel from mean. It is easy to see that n_{ϕ}^{ε} are uncorrelated with their respective $\mathrm{DS}_{\phi}^{\varepsilon}$ terms. We, similar to [12], treat them as worst-case additive Gaussian noise, an approximation which is tight for mMIMO systems [12]. Using (11)-(12), we next derive an achievable SE lower bound.

$$\begin{aligned}
n_k^d = &\underbrace{\tilde{a}\sqrt{\rho_d}\sum_{m\in\mathcal{M}_k^d}\sqrt{\eta_{mk}}\left(\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mk}^d\right)^* - \mathbb{E}\left\{\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mk}^d\right)^*\right\}\right)s_k^d}_{\text{beamforming uncertainty, } \mathrm{BU}_k^d} + \underbrace{\sqrt{\rho_u}\sum_{l=1}^{K_u}h_{kl}\sqrt{\theta_l}s_l^u}_{\mathrm{UDI}_k^d}\\
&+ \underbrace{\tilde{a}\sqrt{\rho_d}\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}\backslash k}\sqrt{\eta_{mq}}\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mq}^d\right)^* s_q^d}_{\mathrm{MUI}_k^d} + \underbrace{\sqrt{\rho_d}\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}}\left(\boldsymbol{g}_{mk}^d\right)^T\left(\hat{\boldsymbol{g}}_{mq}^d\right)^*\varsigma_{mq}^d}_{\mathrm{TQD}_k^d} + w_k^d
\end{aligned}\tag{12}$$

$$\begin{aligned}
n_l^u = &\underbrace{\tilde{a}\sqrt{\rho_u}\sqrt{\theta_l}\sum_{m\in\mathcal{M}_l^u}\left(\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{g}_{ml}^u - \mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{g}_{ml}^u\right\}\right)s_l^u}_{\text{beamforming uncertainty, } \mathrm{BU}_l^u} + \underbrace{\tilde{a}\sum_{m\in\mathcal{M}_l^u}\sum_{q=1,q\neq l}^{K_u}\sqrt{\rho_u}\sqrt{\theta_q}\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{g}_{mq}^u s_q^u}_{\mathrm{MUI}_l^u}\\
&+ \underbrace{\tilde{a}\sqrt{\rho_d}\sum_{m\in\mathcal{M}_l^u}\sum_{i=1}^{M}\sum_{k\in\kappa_{di}}\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\left(\tilde{a}\sqrt{\eta_{ik}}s_k^d + \varsigma_{ik}^d\right)}_{\mathrm{RI}_l^u} + \underbrace{\tilde{a}\sum_{m\in\mathcal{M}_l^u}\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\boldsymbol{w}_m^u}_{N_l^u} + \underbrace{\sum_{m\in\mathcal{M}_l^u}\varsigma_{ml}^u}_{\mathrm{TQD}_l^u}
\end{aligned}\tag{13}$$

Theorem 1: An achievable lower bound to the SE for the kth downlink UE with MRT and the lth uplink UE with MRC can be expressed respectively as

$$S_k^d = \tau_f \log_2\left(1 + \frac{\left(\sum_{m\in\mathcal{M}_k^d} A_{mk}^d\sqrt{\eta_{mk}}\right)^2}{\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}} B_{kmq}^d\eta_{mq} + \sum_{l=1}^{K_u} D_{kl}^d\theta_l + 1}\right), \quad (14)$$

$$S_l^u = \tau_f \log_2\left(1 + \frac{A_l^u\theta_l}{\sum_{q=1}^{K_u} B_{lq}^u\theta_q + \sum_{i=1}^{M}\sum_{k\in\kappa_{di}} D_{lik}^u\eta_{ik} + E_l^u\theta_l + F_l^u}\right), \quad (15)$$

where $\tau_f = (\frac{\tau_c-\tau_t}{\tau_c})$, $A_{mk}^d = \tilde{a}N_t\sqrt{\rho_d}\gamma_{mk}^d$, $B_{kmq}^d = \tilde{b}N_t\rho_d\beta_{mk}^d\gamma_{mq}^d$, $D_{kl}^d = \rho_u\tilde{\beta}_{kl}$, $A_l^u = \tilde{a}^2N_r^2\rho_u(\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u)^2$, $B_{lq}^u = \tilde{b}N_r\rho_u\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u\beta_{mq}^u$, $D_{lik}^u = \tilde{b}^2N_rN_t\rho_d\gamma_{ik}^d\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u\beta_{\mathrm{RI},mi}\gamma_{\mathrm{RI}}$, $E_l^u = (\tilde{b}-\tilde{a}^2)N_r^2\rho_u\sum_{m\in\mathcal{M}_l^u}(\gamma_{ml}^u)^2$, and $F_l^u = \tilde{b}N_r\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u$. Here $\boldsymbol{\eta} \triangleq \{\eta_{mk}\} \in \mathbb{C}^{M\times K_d}$, $\boldsymbol{\Theta} \triangleq \{\theta_l\} \in \mathbb{C}^{K_u\times 1}$ and $\boldsymbol{\nu} \triangleq \{\nu_m\} \in \mathbb{C}^{M\times 1}$ are the variables on which the SE is dependent. We recall from Section II that $\tilde{a}$ and $\tilde{b}$ in (14)-(15) depend on the number of quantization bits, $\boldsymbol{\nu}$.

Proof: Refer to Appendix B. The SE expressions are functions of large scale fading coefficients, γ_{mk}^d and γ_{ml}^u, which we will use to optimize WSEE. *This is unlike [21] which requires instantaneous channel while optimizing SE-GEE metric.* ■

Remark 1: MRC/MRT has tractable SE expression that depend solely on large-scale channel statistics, which remain constant over *hundreds* of coherence intervals [28]. This is in contrast to zero-forcing designs which yield better SE but not tractable SE expressions [2]. Further, MRC/MRT can be implemented in a distributed fashion with low complexity.

IV. TWO-LAYER DECENTRALIZED WSEE OPTIMIZATION FOR FD CF MMIMO

We now devise a decentralized algorithm which maximizes WSEE by calculating the optimal downlink and uplink power control coefficients $\boldsymbol{\eta}^*$ and $\boldsymbol{\Theta}^*$, respectively. We use "two-layered" approach to decompose WSEE maximization into a sequential process with two distinct individual steps, each of which is called a "layer". The first layer simplifies the non-convex WSEE maximization into a successive convex approximation (SCA) setting. Its output is a generalized convex program (GCP) which needs to be solved iteratively for the optimal solution. The second layer optimally solves above GCP, either centrally through standard interior-point approaches or decentrally using ADMM method. The proposed procedure is outlined in Algorithm 2.

Algorithm 2: Two-Layer Decentralized WSEE Maximization Algorithm

1 *AP selection*: Select APs that serve each UE while satisfying limited fronthaul constraints.
2 *SCA framework (first layer)*: Apply a series of transformations and approximations to recast the non-convex WSEE maximization using successive convex approximation (SCA) framework. The output of first layer is a GCP.
3 *Decentralized ADMM approach (second layer)*: Introduce global and local variables to decouple the problem into multiple sub-problems. Each sub-problem is solved at a distributed (or "D") server, whose solutions are coordinated to obtain the global solution at the central (or "C") server. This procedure is implemented using ADMM.

We use $\varepsilon \triangleq \{d, u\}$ for the downlink and uplink, respectively; $\phi \triangleq \{k, l\}$ for the kth downlink UE and lth uplink UE, respectively; and first define the individual EE for each UE as $\mathrm{EE}_\phi^\varepsilon = \frac{B\cdot S_\phi^\varepsilon}{p_\phi^\varepsilon}$ [23], where B is the system bandwidth, and p_ϕ^ε denotes the power consumed by each UE. The fronthaul links consume power for both downlink and uplink transmission. The APs consume power while transmitting data to the downlink UEs, and the uplink UEs consume power while transmitting their data. The power consumed by the system to transmit data to the kth downlink UE and the power consumed by the lth uplink UE are given respectively as [19], [21]

$$p_k^d = P_{\mathrm{fix}} + N_t\rho_dN_0\sum_{m\in\mathcal{M}_k^d}\frac{1}{\alpha_m}\gamma_{mk}^d\eta_{mk} + P_{\mathrm{tc},k}^d, \quad (16)$$

$$p_l^u = P_{\mathrm{fix}} + \rho_uN_0\frac{1}{\alpha_l'}\theta_l + P_{\mathrm{tc},l}^u. \quad (17)$$

Here α_m, α_l' are power amplifier efficiencies at the mth AP and the lth uplink UE respectively [12], N_0 is the noise power and $P_{\mathrm{tc},k}^d, P_{\mathrm{tc},l}^u$ are the powers required to run the transceiver chains at each antenna of the kth downlink UE and the lth uplink UE, respectively. The power consumed by the AP transceiver chains and the fronthaul between APs and CPU:

$$P_{\mathrm{fix}} = \frac{1}{K}\sum_{m=1}^{M}\left(P_{0,m} + (N_t+N_r)P_{\mathrm{tc},m} + P_{\mathrm{ft}}\frac{R_{\mathrm{fh},m}}{C_{\mathrm{fh},m}}\right). \quad (18)$$

Here $P_{\mathrm{tc},m}$ is the power required to run the transceiver chains at each antenna of the mth AP. The fronthaul power consumption for the mth AP has a fixed component, $P_{0,m}$, and a traffic-dependent component, which attains a maximum value of P_{ft} at full capacity $C_{\mathrm{fh},m}$. The term $R_{\mathrm{fh},m}$, given in (7), is the fronthaul data rate of the mth AP. The WSEE is now defined as the weighted sum of EEs of individual UEs [22].

$$\mathrm{WSEE} = \sum_{k=1}^{K_d} w_k^d\mathrm{EE}_k^d + \sum_{l=1}^{K_u} w_l^u\mathrm{EE}_l^u \triangleq B\left(\sum_{k=1}^{K_d} w_k^d\frac{S_k^d}{p_k^d} + \sum_{l=1}^{K_u} w_l^u\frac{S_l^u}{p_l^u}\right),$$

where w_ϕ^ε are weights assigned to the UEs to account for their heterogeneous EE requirements. The WSEE metric can prioritize the EE requirements of individual UEs by assigning them different weights [23], [24]. For example, it could assign a higher weight to a UE that is more energy-scarce. After omitting the constant B from the objective, the WSEE maximization problem can now be formulated as follows

$$\mathbf{P1}: \max_{\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{\nu}} \sum_{k=1}^{K_d} w_k^d\frac{S_k^d(\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{\nu})}{p_k^d(\boldsymbol{\eta},\boldsymbol{\nu})} + \sum_{l=1}^{K_u} w_l^u\frac{S_l^u(\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{\nu})}{p_l^u(\boldsymbol{\Theta},\boldsymbol{\nu})}$$

$$\text{s.t.}\quad S_k^d(\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{\nu}) \geq S_{ok}^d,\ S_l^u(\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{\nu}) \geq S_{ol}^u, \quad (19a)$$

$$R_{\mathrm{fh},m} \leq C_{\mathrm{fh},m},\ (2), (3). \quad (19b)$$

The quality-of-service (QoS) constraints in (19a) guarantee a minimum SE, denoted by the constants S^d_{ok} and S^u_{ol}, for each downlink and uplink UE respectively. The first constraint in (19b) ensures that the fronthaul transmission rate for all APs is within the capacity limit. We observe that the number of quantization bits $\boldsymbol{\nu}$, if included in problem **P1**, will make it a difficult-to-solve integer optimization problem [15], [19], [34]. We therefore solve it to optimize the power control coefficients $\{\boldsymbol{\eta}, \boldsymbol{\Theta}\}$, by fixing $\boldsymbol{\nu}$ such that it satisfies the first constraint in (19b) [15], [19], and numerically investigate $\boldsymbol{\nu}$ in Section V. We reformulate **P1** as follows

$$\begin{aligned}\mathbf{P2}: \max_{\boldsymbol{\eta},\boldsymbol{\Theta}} \; & \sum_{k=1}^{K_d} w^d_k \frac{S^d_k(\boldsymbol{\eta},\boldsymbol{\Theta})}{p^d_k(\boldsymbol{\eta})} + \sum_{l=1}^{K_u} w^u_l \frac{S^u_l(\boldsymbol{\eta},\boldsymbol{\Theta})}{p^u_l(\boldsymbol{\Theta})} \\ \text{s.t.} \; & S^d_k(\boldsymbol{\eta},\boldsymbol{\Theta}) \geq S^d_{ok}, S^u_l(\boldsymbol{\eta},\boldsymbol{\Theta}) \geq S^u_{ol}, \\ & (2),\; (3). \end{aligned} \tag{20}$$

The objective in **P2** is a sum of ratios, each of which is a PC function (concave-over-linear) of power control coefficients $\{\boldsymbol{\eta}, \boldsymbol{\Theta}\}$. It is, therefore, not guaranteed to be a PC function and Dinkelbach's algorithm cannot be applied to maximize it [22]. This makes it a much harder objective to optimize as opposed to the more commonly studied GEE metric, which is a PC function [22] and has been investigated for CF mMIMO systems [18]–[21].

We now maimize WSEE centrally and decentrally using a two-layered approach. The first layer comprises an SCA framework, which formulates a GCP by approximating the non-convex objective and constraints in **P2** as convex. In the second layer, the approximate GCP formed in the nth SCA iteration is either solved centrally or decentrally using ADMM.

Since the approximate GCP obtained in the first layer, due to coupled optimization variables, is not in the standard ADMM form, we introduce their local and global versions. The sub-problems to update local variables are solved independently, and the local variables are coordinated to calculate the global solution [27], [35]. The updation of variables and coordination continues till ADMM converges. The obtained solution is then used to formulate GCP for the $(n+1)$th SCA iteration.

A. SCA FRAMEWORK

We now first linearize the non-convex objective in **P2** using epigraph transformation as [34]

$$\begin{aligned}\mathbf{P3}: \max_{\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{f}^d,\boldsymbol{f}^u} \; & \sum_{k=1}^{K_d} w^d_k f^d_k + \sum_{l=1}^{K_u} w^u_l f^u_l \\ \text{s.t.} \; & f^d_k \leq \frac{S^d_k(\boldsymbol{\eta},\boldsymbol{\Theta})}{p^d_k(\boldsymbol{\eta})}, f^u_l \leq \frac{S^u_l(\boldsymbol{\eta},\boldsymbol{\Theta})}{p^u_l(\boldsymbol{\Theta})}, \\ & (2),\; (3),\; (20). \end{aligned} \tag{21}$$

Here $\boldsymbol{f}^\varepsilon \triangleq [f^\varepsilon_1 \ldots f^\varepsilon_{K_\varepsilon}] \in \mathbb{C}^{K_\varepsilon \times 1}$ are slack variables [34]. To approximate the non-convex constraints in (20) and (21) as convex, we substitute S^d_k and S^u_l from (14)-(15) and cross-multiply the terms p^d_k, p^u_l and f^d_k, f^u_l in (21). We also introduce slack variables $\boldsymbol{\Psi}^\varepsilon \triangleq [\Psi^\varepsilon_1, \ldots, \Psi^\varepsilon_{K_\varepsilon}] \in \mathbb{C}^{K_\varepsilon \times 1}$, $\boldsymbol{\zeta}^\varepsilon \triangleq [\zeta^\varepsilon_1, \ldots, \zeta^\varepsilon_{K_\varepsilon}] \in \mathbb{C}^{K_\varepsilon \times 1}$ and equivalently cast **P3** as [22]

$$\begin{aligned}2\mathbf{P4}: \max_{\substack{\boldsymbol{\eta},\boldsymbol{\Theta},\boldsymbol{f}^d,\boldsymbol{f}^u,\\ \boldsymbol{\Psi}^d,\boldsymbol{\Psi}^u,\boldsymbol{\zeta}^d,\boldsymbol{\zeta}^u}} \; & \sum_{k=1}^{K_d} w^d_k f^d_k + \sum_{l=1}^{K_u} w^u_l f^u_l \\ \text{s.t.}\; p^d_k \leq \frac{(\Psi^d_k)^2}{f^d_k}, p^u_l \leq \frac{\left(\Psi^u_l\right)^2}{f^u_l}, & \qquad (22a) \\ \left(\Psi^d_k\right)^2 \leq \tau_f \log_2\left(1+\zeta^d_k\right), \left(\Psi^u_l\right)^2 \leq \tau_f \log_2\left(1+\zeta^u_l\right), & \qquad (22b) \\ \zeta^d_k \leq \frac{\left(\sum_{m\in\mathcal{M}^d_k} A^d_{mk}\sqrt{\eta_{mk}}\right)^2}{\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}} B^d_{kmq}\eta_{mq} + \sum_{l=1}^{K_u} D^d_{kl}\theta_l + 1}, & \qquad (22c) \\ \zeta^u_l \leq \frac{A^u_l\theta_l}{\sum_{q=1}^{K_u} B^u_{lq}\theta_q + \sum_{i=1}^{M}\sum_{k\in\kappa_{di}} D^u_{lik}\eta_{ik} + E^u_l\theta_l + F^u_l}, & \qquad (22d) \\ \log_2\left(1+\zeta^d_k\right) \geq S^d_{ok}/\tau_f, \log_2\left(1+\zeta^u_l\right) \geq S^u_{ol}/\tau_f, & \\ (2),\; (3). & \qquad (22e) \end{aligned}$$

We introduce the variable $c_{mk} \triangleq \sqrt{\eta_{mk}}$ and denote $\boldsymbol{C} \triangleq \{c_{mk}\} \in \mathbb{C}^{M \times K_d}$ to remove concave terms in (22c) arising due to $\sqrt{\eta_{mk}}$ and facilitate its conversion into a convex constraint. We introduce additional slack variables $\boldsymbol{\lambda}^\varepsilon \triangleq [\lambda^\varepsilon_1, \ldots, \lambda^\varepsilon_{K_\varepsilon}] \in \mathbb{C}^{K_\varepsilon \times 1}$ to further simplify the non-convex constraints (22c)-(22d). We now cast **P4** equivalently as

$$\begin{aligned}\mathbf{P5}: \max_{\substack{\boldsymbol{C},\boldsymbol{\Theta},\boldsymbol{f}^d,\boldsymbol{f}^u\\ \boldsymbol{\Psi}^d,\boldsymbol{\Psi}^u,\boldsymbol{\zeta}^d,\\ \boldsymbol{\zeta}^u,\boldsymbol{\lambda}^d,\boldsymbol{\lambda}^u}} \; & \sum_{k=1}^{K_d} w^d_k f^d_k + \sum_{l=1}^{K_u} w^u_l f^u_l \\ \text{s.t.}\; \sum_{m=1}^{M}\sum_{q\in\kappa_{dm}} B^d_{kmq}c^2_{mq} + \sum_{l=1}^{K_u} D^d_{kl}\theta_l + 1 \leq \frac{\left(\lambda^d_k\right)^2}{\zeta^d_k}, & \qquad (23a) \\ \sum_{q=1}^{K_u} B^u_{lq}\theta_q + \sum_{i=1}^{M}\sum_{k\in\kappa_{di}} D^u_{lik}c^2_{ik} + E^u_l\theta_l + F^u_l \leq \frac{\left(\lambda^u_l\right)^2}{\zeta^u_l}, & \qquad (23b) \\ \lambda^d_k \leq \sum_{m\in\mathcal{M}^d_k} A^d_{mk}c_{mk}, \left(\lambda^u_l\right)^2 \leq A^u_l\theta_l, & \qquad (23c) \\ \lambda^d_k \geq 0, \tilde{b}\sum_{k\in\kappa_{dm}} \gamma^d_{mk}c^2_{mk} \leq \frac{1}{N_t}, c_{mk} \geq 0, & \\ (22a),\; (22b),\; (22e),\; (3). & \qquad (23d) \end{aligned}$$

We note that **P5** has all convex constraints except (22a) and (23a)-(23b). Since a first-order Taylor approximation is a global under-estimator of a convex function [34], we now linearize the right-hand side of these constraints. At the nth iteration, we substitute first-order Taylor approximation $\frac{f_1^2}{f_2} \geq 2\frac{f_1^{(n)}}{f_2^{(n)}}f_1 - \frac{(f_1^{(n)})^2}{(f_2^{(n)})^2}f_2 \triangleq \Lambda^{(n)}(\frac{f_1^2}{f_2})$ and use (16)-(17) to recast **P5** into a GCP:

$$\mathbf{P6}: \max_{\substack{\boldsymbol{C},\boldsymbol{\Theta},\boldsymbol{f}^d,\boldsymbol{f}^u\\ \boldsymbol{\Psi}^d,\boldsymbol{\Psi}^u,\boldsymbol{\zeta}^d,\\ \boldsymbol{\zeta}^u,\boldsymbol{\lambda}^d,\boldsymbol{\lambda}^u}} \; \sum_{k=1}^{K_d} w^d_k f^d_k + \sum_{l=1}^{K_u} w^u_l f^u_l$$

Algorithm 3: Centralized WSEE Maximization Algorithm

```
Input: i) Initialize power control coefficients $\{\boldsymbol{C}, \boldsymbol{\Theta}\}^{(1)}$ by
       allocating equal power to all downlink UEs being served
       and full power to all uplink UEs. Set $n = 1$.
ii) Initialize $\{\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}\}^{(1)}$ by replacing (23c),
(24a)-(24b), (22b) and (24c)-(24d) by equality.
Output: Globally optimal power control coefficients $\{\boldsymbol{C}, \boldsymbol{\Theta}\}^*$
while $\|\boldsymbol{r}_{SCA}^{(n)}\| \leq \epsilon_{SCA}$ do
    Solve P6 for the $n$th SCA iteration to obtain optimal variables,
    $\{\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}, \boldsymbol{C}, \boldsymbol{\Theta}\}^{*,(n)}$.
    Assign the SCA iterates for the $(n+1)$th iteration,
    $\{\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}, \boldsymbol{C}, \boldsymbol{\Theta}\}^{(n+1)} =$
    $\{\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}, \boldsymbol{C}, \boldsymbol{\Theta}\}^{*,(n)}$.
```

$$\text{s.t.} \sum_{m=1}^{M} \sum_{q \in \kappa_{dm}} B_{kmq}^{d} c_{mq}^{2} + \sum_{l=1}^{K_u} D_{kl}^{d} \theta_l + 1 \leq \Lambda^{(n)} \left(\frac{\left(\lambda_k^d\right)^2}{\zeta_k^d} \right), \tag{24a}$$

$$\sum_{q=1}^{K_u} B_{lq}^{u} \theta_q + \sum_{i=1}^{M} \sum_{k \in \kappa_{di}} D_{lik}^{u} c_{ik}^{2} + E_l^u \theta_l + F_l^u \leq \Lambda^{(n)} \left(\frac{\left(\lambda_l^u\right)^2}{\zeta_l^u} \right), \tag{24b}$$

$$P_{\text{fix}} + N_t \rho_d N_0 \sum_{m \in \mathcal{M}_k^d} \frac{\gamma_{mk}^d c_{mk}^2}{\alpha_m} + P_{\text{tc},k}^d \leq \Lambda^{(n)} \left(\frac{\left(\Psi_k^d\right)^2}{f_k^d} \right), \tag{24c}$$

$$P_{\text{fix}} + \rho_u N_0 \frac{\theta_l}{\alpha_l'} + P_{\text{tc},l}^u \leq \Lambda^{(n)} \left(\frac{\left(\Psi_l^u\right)^2}{f_l^u} \right),$$
$$\text{(3), (22b), (22e), (23c), (23d).} \tag{24d}$$

We next provide a centralized SCA to solve **P6** in the second layer in Algorithm 3.

The SCA procedure converges when $\|\boldsymbol{r}_{\text{SCA}}^{(n)}\| = \sqrt{\|\boldsymbol{C}^{(n+1)} - \boldsymbol{C}^{(n)}\|_F^2 + \|\boldsymbol{\Theta}^{(n+1)} - \boldsymbol{\Theta}^{(n)}\|^2}$ has magnitude $\|\boldsymbol{r}_{\text{SCA}}^{(n)}\| \leq \epsilon_{\text{SCA}}$, where ϵ_{SCA} is the convergence threshold.

Remark 2 Convergence of Centralized Algorithm: At the nth SCA iteration, **P6** is obtained from **P5** by applying first-order Taylor approximations to the constraints (22a) and (23a)-(23b). These approximations are of the form $\Lambda(\boldsymbol{x}) \triangleq \frac{x_1^2}{x_2} \geq 2\frac{x_1^{(n)}}{x_2^{(n)}} x_1 - (\frac{x_1^{(n)}}{x_2^{(n)}})^2 x_2 \triangleq \bar{\Lambda}(\boldsymbol{x}, \boldsymbol{x}^{(n)})$. It is easy to show that **P6** is the *inner-approximation problem* for **P5**, where we replace each of the constraints (22a) and (23a)-(23b), denoted here as $g_i(\boldsymbol{x}) \leq 0$, $i = 1, 2, 3$, with a convex approximation of the form $\bar{g}_i(\boldsymbol{x}, \boldsymbol{x}^{(n)}) \leq 0$, $i = 1, 2, 3$. For each of the approximations, it can be easily shown that the following properties hold [36]: i) $g_i(\boldsymbol{x}) \leq \bar{g}_i(\boldsymbol{x}, \boldsymbol{x}^{(n)})$ for all feasible $\boldsymbol{x}$; ii) $g_i(\boldsymbol{x}^{(n)}) = \bar{g}_i(\boldsymbol{x}^n, \boldsymbol{x}^{(n)})$; and $\frac{\partial g_i(\boldsymbol{x}^{(n)})}{\partial x_j} = \frac{\partial \bar{g}_i(\boldsymbol{x}^n, \boldsymbol{x}^{(n)})}{\partial x_j}, j = 1, 2$. The constraints in **P6** also satisfy Slater's conditions [34].

This implies that Algorithm 3, by solving the inner-approximation problem, always converges to a KKT point of **P2** due to [36]. It must be noted here that even though Algorithm 3 solves the approximate problem **P6** in each SCA iteration, it is provably optimal after sufficient number of iterations. This is due to the fact that it provably converges to a KKT point of **P2** which is an optimal solution [34].

B. DECENTRALIZED ADMM APPROACH

We now use ADMM to decentrally solve **P6** in the second layer, an approach well-suited for CPUs with multiple distributed D-servers, connected via a central C-server [25], [26]. The ADMM method decomposes a central problem into multiple sub-problems, each of which is solved by a D-server locally and independently. The C-server combines the local solutions to obtain a global solution. We observe that the constraints in (24a)-(24b) couple the power control coefficients of different uplink and downlink UEs. We next introduce global variables for the power control coefficients at the C-server, with local copies at the D-servers to decouple **P6** into sub-problems for each UE. We observe that the constraints in **P6** for the downlink and uplink UEs can be divided between downlink and uplink D-servers, respectively. The D-servers solve sub-problems defined for each downlink and uplink UE. We first define local feasible sets at the nth SCA iteration for them, which are denoted as $\mathcal{S}_k^{d,(n)}$ and $\mathcal{S}_l^{u,(n)}$, respectively. These sets are given as follows

$$\mathcal{S}_k^{d,(n)} = \left\{ f_k^d, \Psi_k^d, \zeta_k^d, \lambda_k^d, \widetilde{\boldsymbol{C}}_k^d, \widetilde{\boldsymbol{\Theta}}_k^d \middle| \tilde{b} \sum_{q \in \kappa_{dm}} \gamma_{mk}^d \left(\tilde{c}_{mq,k}^d \right)^2 \leq \frac{1}{N_t}, \right. \tag{25a}$$

$$\lambda_k^d \leq \sum_{m \in \mathcal{M}_k^d} A_{mk}^d \tilde{c}_{mk,k}^d, \sum_{m=1}^{M} \sum_{q \in \kappa_{dm}} B_{kmq}^d \left(\tilde{c}_{mq,k}^d \right)^2$$
$$+ \sum_{l=1}^{K_u} D_{kl}^d \tilde{\theta}_{l,k}^d + 1 \leq \Lambda^{(n)} \left(\frac{\left(\lambda_k^d\right)^2}{\zeta_k^d} \right), \tag{25b}$$

$$\left(\Psi_k^d\right)^2 \leq \tau_f \log_2\left(1 + \zeta_k^d\right), P_{\text{fix}} + N_t \rho_d N_0 \sum_{m \in \mathcal{M}_k^d} \frac{1}{\alpha_m} \gamma_{mk}^d \left(\tilde{c}_{mk,k}^d \right)^2$$
$$+ P_{\text{tc},k}^d \leq \Lambda^{(n)} \left(\frac{\left(\Psi_k^d\right)^2}{f_k^d} \right), \tag{25c}$$

$$\tilde{c}_{mq,k}^d \geq 0 \, \forall q = 1 \text{ to } K_d, 0 \leq \tilde{\theta}_{l,k}^d \leq 1, \lambda_k^d \geq 0,$$
$$\left. \log_2\left(1 + \zeta_k^d\right) \geq S_{ok}^d / \tau_f \right\}, \tag{25d}$$

$$\mathcal{S}_l^{u,(n)} = \left\{ f_l^u, \Psi_l^u, \zeta_l^u, \lambda_l^u, \widetilde{\boldsymbol{C}}_{\boldsymbol{l}}^{\boldsymbol{u}}, \widetilde{\boldsymbol{\Theta}}_{\boldsymbol{l}}^{\boldsymbol{u}} \middle| \tilde{b} \sum_{k \in \kappa_{dm}} \gamma_{mk}^d (\tilde{c}_{mk,l}^u)^2 \leq \frac{1}{N_t}, \right. \tag{25e}$$

$$\left(\lambda_l^u\right)^2 \leq A_l^u \tilde{\theta}_{l,l}^u, \sum_{q=1}^{K_u} B_{lq}^u \tilde{\theta}_{q,l}^u + \sum_{i=1}^{M} \sum_{k \in \kappa_{di}} D_{lik}^u (\tilde{c}_{ik,l}^u)^2$$
$$+ E_l^u \tilde{\theta}_{l,l}^u + F_l^u \leq \Lambda^{(n)} \left(\frac{\left(\lambda_l^u\right)^2}{\zeta_l^u} \right), \tag{25f}$$

$$\left(\Psi_l^u\right)^2 \leq \tau_f \log_2(1 + \zeta_l^u),$$
$$P_{\text{fix}} + \rho_u N_0 \frac{1}{\alpha_l'} \tilde{\theta}_{l,l}^u + P_{\text{tc},l}^u \leq \Lambda^{(n)} \left(\frac{\left(\Psi_l^u\right)^2}{f_l^u} \right), \tag{25g}$$

$$\tilde{c}_{mk,l}^u \geq 0, 0 \leq \tilde{\theta}_{q,l}^u \leq 1 \, \forall q = 1 \text{ to } K_u,$$
$$\left. \log_2(1 + \zeta_l^u) \geq S_{ol}^u / \tau_f \right\}. \tag{25h}$$

Here $\widetilde{\boldsymbol{C}}_k^d, \widetilde{\boldsymbol{C}}_l^u \in \mathbb{C}^{M\times K_d}$ and $\widetilde{\boldsymbol{\Theta}}_k^d, \widetilde{\boldsymbol{\Theta}}_l^u \in \mathbb{C}^{K_u\times 1}$ are local copies at the D-server of the corresponding global variables at the C-server, which are denoted as $\widetilde{\boldsymbol{C}} \in \mathbb{C}^{M\times K_d}$ and $\widetilde{\boldsymbol{\Theta}} \in \mathbb{C}^{K_u\times 1}$ respectively, and represent the downlink and uplink power control coefficients, $\boldsymbol{C}$ and $\boldsymbol{\Theta}$, in **P6**. We note that each D-server has its local power control variables and hence the constraints in (25), which are all convex, are independent for each D-server. This ensures that the sets $\mathcal{S}_k^{d,(n)}$ and $\mathcal{S}_l^{u,(n)}$ are convex. We define the sets of local variables for the D-servers corresponding to the downlink and uplink UEs as $\boldsymbol{\Omega}_k^d \triangleq [\widetilde{\boldsymbol{C}}_k^d, \widetilde{\boldsymbol{\Theta}}_k^d, f_k^d, \Psi_k^d, \lambda_k^d, \zeta_k^d]$ and $\boldsymbol{\Omega}_l^u \triangleq [\widetilde{\boldsymbol{C}}_l^u, \widetilde{\boldsymbol{\Theta}}_l^u, f_l^u, \Psi_l^u, \lambda_l^u, \zeta_l^u]$ respectively.

We now reformulate **P6** as follows

$$\begin{aligned}
\textbf{2P7}: \max_{\widetilde{\boldsymbol{C}},\widetilde{\boldsymbol{\Theta}},\boldsymbol{\Omega}_k^d,\boldsymbol{\Omega}_l^u} \quad & \sum_{k=1}^{K_d} w_k^d f_k^d + \sum_{l=1}^{K_u} w_l^u f_l^u \\
\text{s.t.} \quad & \boldsymbol{\Omega}_k^d \in \mathcal{S}_k^{d,(n)}, \boldsymbol{\Omega}_l^u \in \mathcal{S}_l^{u,(n)}, && (26a) \\
& \widetilde{\boldsymbol{C}}_k^d = \widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{C}}_l^u = \widetilde{\boldsymbol{C}}, && (26b) \\
& \widetilde{\boldsymbol{\Theta}}_k^d = \widetilde{\boldsymbol{\Theta}}, \widetilde{\boldsymbol{\Theta}}_l^u = \widetilde{\boldsymbol{\Theta}}. && (26c)
\end{aligned}$$

To ensure that the global variables at the C-server have identical local copies maintained at the D-servers, we introduce the consensus constraints (26b)-(26c). The ADMM algorithm can now be readily applied to **P7** as it is in the global consensus form [35].

We use $\varepsilon \triangleq \{d, u\}$ to denote the downlink and uplink respectively, and $\phi \triangleq \{k, l\}$ to denote kth the downlink UE and lth uplink UE, respectively. The sub-problems of individual D-servers can now be written as follows

$$\begin{aligned}
\textbf{P7b}: \max_{\widetilde{\boldsymbol{C}},\widetilde{\boldsymbol{\Theta}},\boldsymbol{\Omega}_\phi^\varepsilon} \quad & w_\phi^\varepsilon f_\phi^\varepsilon \\
\text{s.t.} \quad & \boldsymbol{\Omega}_\phi^\varepsilon \in \mathcal{S}_\phi^{\varepsilon,(n)}, \widetilde{\boldsymbol{C}}_\phi^\varepsilon = \widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}_\phi^\varepsilon = \widetilde{\boldsymbol{\Theta}}.
\end{aligned}$$

We now define auxiliary functions for the objective in **P7b** as follows

$$q_\phi^\varepsilon\left(\boldsymbol{\Omega}_\phi^\varepsilon\right) \triangleq \begin{cases} w_\phi^\varepsilon f_\phi^\varepsilon, & \boldsymbol{\Omega}_\phi^\varepsilon \in \mathcal{S}_\phi^{\varepsilon,(n)}, \\ -\infty, & \text{otherwise.} \end{cases} \tag{27}$$

We write, using (27), the augmented Lagrangian function for **P7** as

$$\begin{aligned}
&\mathcal{L}^{(n)}\left(\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}, \left\{\boldsymbol{\Omega}_k^d, \boldsymbol{\chi}_k^d, \boldsymbol{\xi}_k^d\right\}, \left\{\boldsymbol{\Omega}_l^u, \boldsymbol{\chi}_l^u, \boldsymbol{\xi}_l^u\right\}\right) \\
&= \sum_{k=1}^{K_d}\Bigg(q_k^d\left(\boldsymbol{\Omega}_k^d\right) - \langle \boldsymbol{\chi}_k^d, \widetilde{\boldsymbol{C}}_k^d - \widetilde{\boldsymbol{C}}\rangle - \frac{\rho_C}{2}\left\|\widetilde{\boldsymbol{C}}_k^d - \widetilde{\boldsymbol{C}}\right\|_F^2 \\
&\quad -\left\langle \boldsymbol{\xi}_k^d, \widetilde{\boldsymbol{\Theta}}_k^d - \widetilde{\boldsymbol{\Theta}}\right\rangle - \frac{\rho_\theta}{2}\left\|\widetilde{\boldsymbol{\Theta}}_k^d - \widetilde{\boldsymbol{\Theta}}\right\|^2\Bigg) \\
&\quad + \sum_{l=1}^{K_u}\Bigg(q_l^u(\boldsymbol{\Omega}_l^u) - \left\langle \boldsymbol{\chi}_l^u, \widetilde{\boldsymbol{C}}_l^u - \widetilde{\boldsymbol{C}}\right\rangle - \frac{\rho_C}{2}\left\|\widetilde{\boldsymbol{C}}_l^u - \widetilde{\boldsymbol{C}}\right\|_F^2 \\
&\quad -\left\langle \boldsymbol{\xi}_l^u, \widetilde{\boldsymbol{\Theta}}_l^u - \widetilde{\boldsymbol{\Theta}}\right\rangle - \frac{\rho_\theta}{2}\left\|\widetilde{\boldsymbol{\Theta}}_l^u - \widetilde{\boldsymbol{\Theta}}\right\|^2\Bigg),
\end{aligned} \tag{28}$$

where $\rho_C, \rho_\theta > 0$ are the penalty parameters corresponding to the global variables $\widetilde{\boldsymbol{C}}$ and $\widetilde{\boldsymbol{\Theta}}$ respectively, and $\boldsymbol{\chi}_\phi^\varepsilon \in \mathbb{C}^{M\times K_d}$, $\boldsymbol{\xi}_\phi^\varepsilon \in \mathbb{C}^{K_u\times 1}$ are the Lagrangian variables associated with the equality constraints (26b) and (26c), respectively. The quadratic penalty terms are added to the objective to penalise equality constraints violations, and to enable the ADMM to converge by relaxing constraints of finiteness and strict convexity [35].

We note that the augmented Lagrangian in (28) is not decomposable in general for the problem formulation in **P7b** [34]. The auxiliary functions defined in (27) enable us to decompose it and formulate sub-problems for the D-servers. In ADMM method, the D-servers independently solve the sub-problems and update the local variables, which are collected by the C-server to update the global variables [35]. In the $(p+1)$th iteration, following steps are executed in succession.

1) *Local Computation:* The D-servers for each UE solve **P8** to update the local variables as

$$\begin{aligned}
\textbf{P8}: \boldsymbol{\Omega}_\phi^{\varepsilon,(p+1)} = \arg\max_{\boldsymbol{\Omega}_\phi^\varepsilon} \quad & q_\phi^\varepsilon\left(\boldsymbol{\Omega}_\phi^\varepsilon\right) - \left\langle \boldsymbol{\chi}_\phi^{\varepsilon,(p)}, \widetilde{\boldsymbol{C}}_\phi^\varepsilon - \widetilde{\boldsymbol{C}}^{(p)}\right\rangle \\
& - \left\langle \boldsymbol{\xi}_\phi^{\varepsilon,(p)}, \widetilde{\boldsymbol{\Theta}}_\phi^\varepsilon - \widetilde{\boldsymbol{\Theta}}^{(p)}\right\rangle \\
& - \frac{\rho_C^{(p)}}{2}\left\|\widetilde{\boldsymbol{C}}_\phi^\varepsilon - \widetilde{\boldsymbol{C}}^{(p)}\right\|_F^2 - \frac{\rho_\theta^{(p)}}{2}\left\|\widetilde{\boldsymbol{\Theta}}_\phi^\varepsilon - \widetilde{\boldsymbol{\Theta}}^{(p)}\right\|^2.
\end{aligned} \tag{29}$$

2) *Lagrangian Multipliers Update:* The D-servers now update the Lagrangian multipliers as

$$\boldsymbol{\chi}_\phi^{\varepsilon,(p+1)} = \boldsymbol{\chi}_\phi^{\varepsilon,(p)} + \rho_C^{(p)}\left(\widetilde{\boldsymbol{C}}_\phi^{\varepsilon,(p+1)} - \widetilde{\boldsymbol{C}}^{(p)}\right) \tag{30}$$

$$\boldsymbol{\xi}_\phi^{\varepsilon,(p+1)} = \boldsymbol{\xi}_\phi^{\varepsilon,(p)} + \rho_\theta^{(p)}\left(\widetilde{\boldsymbol{\Theta}}_\phi^{\varepsilon,(p+1)} - \widetilde{\boldsymbol{\Theta}}^{(p)}\right). \tag{31}$$

3) *Global Aggregation and Computation:* The C-server now collects the updated local variables and Lagrangian multipliers from the D-servers and updates the global variables $\{\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}\}$.

$$\begin{aligned}
\textbf{P9}: \left\{\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}\right\}^{(p+1)} = \arg\max_{\widetilde{\boldsymbol{C}},\widetilde{\boldsymbol{\Theta}}} \quad & \mathcal{L}^{(n)}\Big(\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}, \\
& \left\{\boldsymbol{\Omega}_k^d, \boldsymbol{\chi}_k^d, \boldsymbol{\xi}_k^d\right\}^{(p+1)}, \left\{\boldsymbol{\Omega}_l^u, \boldsymbol{\chi}_l^u, \boldsymbol{\xi}_l^u\right\}^{(p+1)}\Big).
\end{aligned}$$

Using (28) and maximizing w.r.t. each global variable, we obtain a closed form solution

$$\begin{aligned}
\widetilde{\boldsymbol{C}}^{(p+1)} = \frac{1}{K}\Bigg(&\sum_{k=1}^{K_d}\left[\widetilde{\boldsymbol{C}}_k^{d,(p+1)} + \frac{1}{\rho_C^{(p)}}\boldsymbol{\chi}_k^{d,(p+1)}\right] \\
&+ \sum_{l=1}^{K_u}\left[\widetilde{\boldsymbol{C}}_l^{u,(p+1)} + \frac{1}{\rho_C^{(p)}}\boldsymbol{\chi}_l^{u,(p+1)}\right]\Bigg),
\end{aligned} \tag{32}$$

$$\begin{aligned}
\widetilde{\boldsymbol{\Theta}}^{(p+1)} = \frac{1}{K}\Bigg(&\sum_{k=1}^{K_d}\left[\widetilde{\boldsymbol{\Theta}}_k^{d,(p+1)} + \frac{1}{\rho_\theta^{(p)}}\boldsymbol{\xi}_k^{d,(p+1)}\right] \\
&+ \sum_{l=1}^{K_u}\left[\widetilde{\boldsymbol{\Theta}}_l^{u,(p+1)} + \frac{1}{\rho_\theta^{(p)}}\boldsymbol{\xi}_l^{u,(p+1)}\right]\Bigg).
\end{aligned} \tag{33}$$

The updated global variables in (32)-(33) are broadcasted by the C-server to all the D-servers.

4) *Residue Calculation and Penalty Parameter Updates:* The C-server calculates the squared magnitude of the primal and dual residuals, denoted as $\boldsymbol{r}_{\text{ADMM}}$ and $\boldsymbol{s}_{\text{ADMM}}$ respectively, as [35]

$$\left\|\boldsymbol{r}_{\text{ADMM}}^{(p+1)}\right\|_2^2 = \sum_{k=1}^{K_d}\left(\left\|\widetilde{\boldsymbol{C}}_k^d - \widetilde{\boldsymbol{C}}\right\|_F^2 + \left\|\widetilde{\boldsymbol{\Theta}}_k^d - \widetilde{\boldsymbol{\Theta}}\right\|_2^2\right)^{(p+1)} + \sum_{l=1}^{K_u}\left(\left\|\widetilde{\boldsymbol{C}}_l^u - \widetilde{\boldsymbol{C}}\right\|_F^2 + \left\|\widetilde{\boldsymbol{\Theta}}_l^u - \widetilde{\boldsymbol{\Theta}}\right\|_2^2\right)^{(p+1)}, \quad (34)$$

$$\left\|\boldsymbol{s}_{\text{ADMM}}^{(p+1)}\right\|_2^2 = K\left(\left\|\widetilde{\boldsymbol{C}}^{(p+1)} - \widetilde{\boldsymbol{C}}^{(p)}\right\|_F^2 + \left\|\widetilde{\boldsymbol{\Theta}}^{(p+1)} - \widetilde{\boldsymbol{\Theta}}^{(p)}\right\|_2^2\right). \quad (35)$$

The C-server now compares the primal and dual residual norms obtained in (34)-(35). To accelerate convergence, it updates the penalty parameters for the $(p+1)$th ADMM iteration, $\rho_{\{C\}}^{(p+1)}$ and $\rho_{\{\theta\}}^{(p+1)}$, appropriately as follows [37]:

$$\rho_{\{C,\theta\}}^{(p+1)} = \begin{cases} \rho_{\{C,\theta\}}^{(p)}\vartheta^{\text{incr}}, & \|r^{(p+1)}\|_2 > \mu\|s^{(p+1)}\|_2, \\ \rho_{\{C,\theta\}}^{(p)}/\vartheta^{\text{decr}}, & \|s^{(p+1)}\|_2 > \mu\|r^{(p+1)}\|_2, \\ \rho_{\{C,\theta\}}^{(p)}, & \text{otherwise}. \end{cases} \quad (36)$$

The parameters $\mu > 1, \vartheta^{\text{incr}} > 1, \vartheta^{\text{decr}} > 1$ are tuned to obtain good convergence [37].

Initialization for ADMM: At the $(n+1)$th SCA iteration, we initialize the global variables at the C-server and their local copies at the D-servers with the SCA iteration variables as

$$\tilde{c}_{mk}^{(1)} = c_{mk}^{(n+1)}, \tilde{\theta}_l^{(1)} = \theta_l^{(n+1)}, \widetilde{\boldsymbol{C}}_k^{d,(1)} = \widetilde{\boldsymbol{C}}_l^{u,(1)} = \widetilde{\boldsymbol{C}}^{(1)}, \widetilde{\boldsymbol{\Theta}}_k^{d,(1)} = \widetilde{\boldsymbol{\Theta}}_l^{u,(1)} = \widetilde{\boldsymbol{\Theta}}^{(1)}. \quad (37)$$

ADMM Convergence Criterion: The ADMM can be said to have converged at iteration P if the primal residue is within a pre-determined tolerance limit ϵ_{ADMM}, i.e., $\|r^{(P)}\|_2 \leq \epsilon_{\text{ADMM}}$. The steps (29), (30)-(31), (32)-(33) and (36) are iterated until convergence, after which we obtain the locally optimal power control coefficients $\{\widetilde{\boldsymbol{C}}^*, \widetilde{\boldsymbol{\Theta}}^*\}$. We assign them to the iterates for the $(n+1)$th SCA iteration, i.e., $\boldsymbol{C}^{(n+1)} = \widetilde{\boldsymbol{C}}^*, \boldsymbol{\Theta}^{(n+1)} = \widetilde{\boldsymbol{\Theta}}^*$. This concludes the nth SCA iteration. The SCA is iterated till convergence. The steps for the decentralized WSEE maximization using SCA and ADMM are summarized in Algorithm 4.

Remark 3 Convergence of Proposed Decentralized Algorithm: Algorithm 4 uses the iterative SCA technique with each SCA iteration involving ADMM. The algorithm is thus guaranteed to converge if both SCA and ADMM converge. It must be noted here that Algorithm 4, despite solving an approximate problem **P7** in each ADMM iteration, indeed converges to an optimal solution of the original problem **P2**. This is explained as follows. For a given SCA iteration, the convergence of ADMM is guaranteed and investigated in detail in [35]. Hence, every SCA iteration converges to an optimal solution of the approximate problem **P6**. As discussed in Remark 3, the SCA iterative procedure provably converges to a KKT point of **P2** which is an optimal solution [34].

Algorithm 4: Decentralized WSEE Maximization Algorithm Using SCA and ADMM

```
Input: i) Initialize power control coefficients for SCA, ${\boldsymbol{C}, \boldsymbol{\Theta}}^{(1)}$
       by allocating equal power to downlink UEs and maximum
       power to uplink UEs. Set $n = 1$. Initialize
       ${\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}}^{(1)}$ by replacing inequalities
       (23c), (24a)-(24b), (22b) and (24c)-(24d) by equality, in turn.
Output: Globally optimal power control coefficients ${\boldsymbol{C}, \boldsymbol{\Theta}}^*$
while $\|\boldsymbol{r}_{SCA}\| \leq \epsilon_{SCA}$ do
    Set $p = 1$. Initialize global variables at C-server, ${\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}}^{(1)}$, and
    local variables at D-servers, $\boldsymbol{\Omega}_\phi^{\varepsilon,(1)}$, using (37) and replacing
    inequalities (25b)-(25c) and (25f)-(25g) by equality.
    while $\|\boldsymbol{r}_{ADMM}\| \leq \epsilon_{ADMM}$ do
        Substitute ${\boldsymbol{C}, \boldsymbol{\Theta}, \boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}}^{(n)}$ in
        (25) to obtain feasible sets $\mathcal{S}_\phi^{\varepsilon,(n)}$.
        Solve **P8** at respective D-servers to update local variables
        $\boldsymbol{\Omega}_\phi^{\varepsilon,(p+1)}$.
        Solve (30)-(31) at respective D-servers to update
        Lagrangian multipliers ${\boldsymbol{\chi}, \boldsymbol{\xi}}_\phi^{\varepsilon,(p+1)}$.
        At the C-server, collect the local variables ${\widetilde{\boldsymbol{C}}, \widetilde{\boldsymbol{\Theta}}}_\phi^{\varepsilon,(p+1)}$,
        and the Lagrangian multipliers, ${\boldsymbol{\chi}, \boldsymbol{\xi}}_\phi^{\varepsilon,(p+1)}$, from the
        D-servers and solve (32)-(33) to update the global variables
        $\widetilde{\boldsymbol{C}}^{(p+1)}, \widetilde{\boldsymbol{\Theta}}^{(p+1)}$.
        At the C-server, update penalty parameters $\rho_{C,\theta}^{(p+1)}$
        according to (36) and broadcast them to all D-servers.
    Update $\boldsymbol{C}^{(n+1)} = \widetilde{\boldsymbol{C}}^*, \boldsymbol{\Theta}^{(n+1)} = \widetilde{\boldsymbol{\Theta}}^*$ and obtain
    ${\boldsymbol{f^d}, \boldsymbol{f^u}, \boldsymbol{\lambda^d}, \boldsymbol{\lambda^u}, \boldsymbol{\Psi^d}, \boldsymbol{\Psi^u}, \boldsymbol{\zeta^d}, \boldsymbol{\zeta^u}}^{(n+1)}$ by replacing the
    inequalities (23c), (24a)-(24b), (22b) and (24c)-(24d) by equality.
return ${\boldsymbol{C}, \boldsymbol{\Theta}}^*$.
```

Remark 4 Implementability: The maximal ratio combiner/beamformer considered herein is the simplest receiver/transmitter for a distributed cell-free mMIMO system [2]. Further, the power optimization algorithms require only long-term fading channel coefficients, which remain constant for hundreds of coherence intervals [28]. This is in contrast to the existing work in SE-GEE maximization of FD cell-free massive MIMO systems in [21], which requires instantaneous channel. The current optimization problem whose reduced complexity is discussed below, therefore, needs to be solved over a relaxed time frame, which makes it easily implementable.

C. COMPUTATIONAL COMPLEXITY OF CENTRALIZED AND DECENTRALIZED ALGORITHMS

Before beginning this study, it is worth noting that both centralized Algorithm 3 and decentralized Algorithm 4 comprise of multiple steps that involve solving simple closed form expressions. These steps consume much lesser time than the ones which solve a GCP, typically using interior points methods [34]. We therefore compare the per-iteration complexity of centralized and decentralized algorithms by calculating the complexity of solving the respective GCPs.

- Algorithm 3 solves **P6** in step-1 of each SCA iteration, which has $4(K_u + K_d) + K_u + MK_d$ real variables and $6(K_u + K_d) + M + MK_d$ linear constraints. It has a worst-case computational complexity $\mathcal{O}((10(K_u + K_d) + K_u + M + 2MK_d)^{3/2}(4(K_u + K_d) + K_u + MK_d)^2)$ [38].

- Algorithm 4, in step-2 of each ADMM iteration, solves **P8** at the D-servers *in parallel* to update the local variables. We, therefore, need to analyse the computational complexity at *any one of the* D-servers. Since the downlink has an additional constraint (second one in (25d)), we consider a downlink D-server for worst-case complexity analysis, which in **P8** has $MK_d + K_u + 4$ real variables and $MK_d + M + K_u + 6$ linear constraints. It will have a worst-case computational complexity [38]: $\mathcal{O}((2MK_d + M + 2K_u + 10)^{3/2}(MK_d + K_u + 4)^2)$.

We consider $K_d = K_u = K/2$ uplink and downlink UEs for this analysis. We observe that for a large K, Algorithm 4 has a much lower computational complexity than Algorithm 3.

V. SIMULATION RESULTS

We now numerically investigate the SE and WSEE of a FD CF mMIMO system with limited-capacity fronthaul links. We assume a realistic system model wherein the M APs, K_d downlink UEs and K_u uplink UEs are all scattered randomly in a square of size D km × D km. To avoid the boundary effects [3], we wrap the APs and UEs around the edges [12]. We use $\varepsilon \triangleq \{d, u\}$ to denote downlink and uplink respectively, and $\phi \triangleq \{k, l\}$ to denote kth downlink UE and lth uplink UE, respectively. The large-scale fading coefficients, $\beta_{m\phi}^{\varepsilon}$, are modeled as [18]

$$\beta_{m\phi}^{\varepsilon} = 10^{\frac{\text{PL}_{m\phi}^{\varepsilon}}{10}} 10^{\frac{\sigma_{\text{sd}} z_{m\phi}^{\varepsilon}}{10}}. \tag{38}$$

Here $10^{\frac{\sigma_{\text{sd}} z_{m\phi}^{\varepsilon}}{10}}$ is the log-normal shadowing factor with a standard deviation σ_{sd} (in dB) and $z_{m\phi}^{\varepsilon}$ follows a two-components correlated model [3]. The path loss $\text{PL}_{m\phi}^{\varepsilon}$ (in dB) follows a three-slope model [3], [12].

We, similar to [12], model the large-scale fading coefficients for the inter-AP RI channels, i.e., $\beta_{\text{RI},mi}$, $\forall i \neq m$, as in (38), and assume that the large-scale fading for the intra-AP RI channels, which do not experience shadowing, are modeled as $\beta_{\text{RI},mm} = 10^{\frac{\text{PL}_{\text{RI}}(\text{dB})}{10}}$. The inter-UE large scale fading coefficients, $\tilde{\beta}_{kl}$, are also modeled similar to (38). We consider, for brevity, the same number of quantization bits ν, and the same fronthaul capacity C_{fh} for all links. We, henceforth, denote the transmit powers on the downlink and uplink as p_d $(= \rho_d N_0)$ and p_u $(= \rho_u N_0)$, respectively, and the pilot transmit power as $p_t (= \rho_t N_0)$. We fix the system model values and power consumption model parameters, unless mentioned otherwise, as given in Table 1. These values are commonly used in the literature, e.g., [3], [12], [15].

Validation of SE expressions: We consider an FD CF mMIMO system with i) $M = \{16, 32\}$ APs, each having $N_t = N_r = 8$ transmit and receive antennas, $K_d = 12$ downlink UEs and $K_u = 8$ uplink UEs; and ii) unequal uplink and downlink transmit power, i.e., $p_d = 2p_u = p$. We verify in Fig. 2 the tightness of the SE lower bound derived in (14)-(15), labeled as LB, by comparing it with the numerically-obtained ergodic SE in (10), labeled as upper-bound (UB) as it requires instantaneous CSI. The large-scale fading coefficients are set according to a practical FD CF channel model with parameters specified in Table 1. We, similar to [3], [18], allocate equal power to all downlink UEs and full power to all uplink UEs, i.e., $\eta_{mk} = (bN_t(\sum_{k\in\kappa_{dm}} \gamma_{mk}^{d}))^{-1}, \forall k \in \kappa_{dm}$ and $\theta_l = 1$. *We see that the derived lower bound is tight for both values of M.*

TABLE 1. **Full-duplex cell-free mMIMO system model and power consumption model parameters.**

Parameter	Value
Coverage area side length, D	1 km
Shadowing parameters σ_{sd}, δ	2 dB, 0.5
Bandwidth B	20 Mhz
Length of coherence period, τ_c, coherence time T_c	200 symbols, 1 ms
Fronthaul parameters ν, C_{fh}	2, 10 Mbps
RI parameters γ_{RI}, PL_{RI} (in dB)	−20, −81.1846
AP power parameters, $P_{\text{ft}}, P_{0,m}, P_{\text{tc},m}$ (in W)	10, 0.825, 0.2
UE power parameters, $P_{\text{tc},k}^{d} = P_{\text{tc},l}^{u}$	0.2 W
Pilot power p_t, Noise power N_0	0.2 W, -121.4 dB
Power amplifier efficiencies, α_m, α_l'	0.39, 0.3

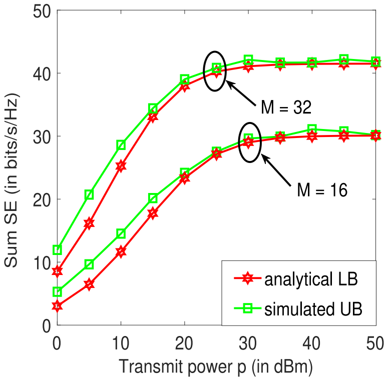
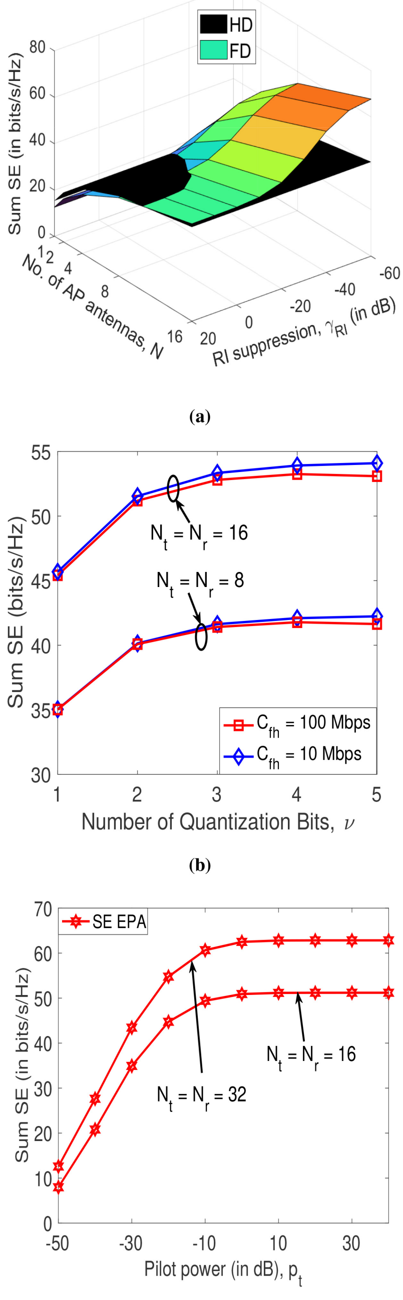


FIGURE 2. **Sum SE vs transmit power, with $N_t = N_r = 8$, $K_d = 12$, $K_u = 8$.**

Sum SE - FD and HD comparison: We consider an FD CF mMIMO system with $M = 32$ APs, $K_d = 12$ downlink UEs, $K_u = 8$ uplink UEs and with transmit powers $p_d = 30$ dBm, $p_u = 27$ dBm on the downlink and uplink. We compare in Fig. 3(a) the FD CF mMIMO system with varying levels of RI suppression factor γ_{RI} and an equivalent HD system which serves uplink and downlink UEs in time-division duplex mode. For the HD system, we i) set $\gamma_{\text{RI}} = 0$ and inter-UE channel gains $\tilde{\beta}_{kl} = 0$; ii) use all AP antennas, i.e., $N = (N_t + N_r)$, during uplink and downlink transmission; and iii) multiply sum SE with a factor of 1/2. We see that the FD system has a significantly higher sum SE than an equivalent HD system, provided the RI suppression is good, i.e., $\gamma_{\text{RI}} \leq -10$ dB. It is important to reemphasize here that the gains in sum SE achieved by the FD transmissions completely vanish with poor RI suppression, i.e., $\gamma_{\text{RI}} > -10$ dB. Moreover we note that, contrary to intuitive expectations, the sum SE does not double, even with significant RI suppression $\gamma_{\text{RI}} \leq -40$ dB. This is due to the UDI experienced by the downlink UEs in a FD CF mMIMO system as shown in Fig. 1, which cannot be mitigated by RI suppression at APs.

Sum SE - variation with quantization bits: We plot in Fig. 3(b) the sum SE by varying the number of fronthaul



(a)

(b)

(c)

FIGURE 3. Sum SE vs a) RI suppression levels, b) Number of quantization bits, and c) pilot power with $M = 32$, $K_d = 12$, $K_u = 8$, $p_d = 2p_u = 30$ dBm.

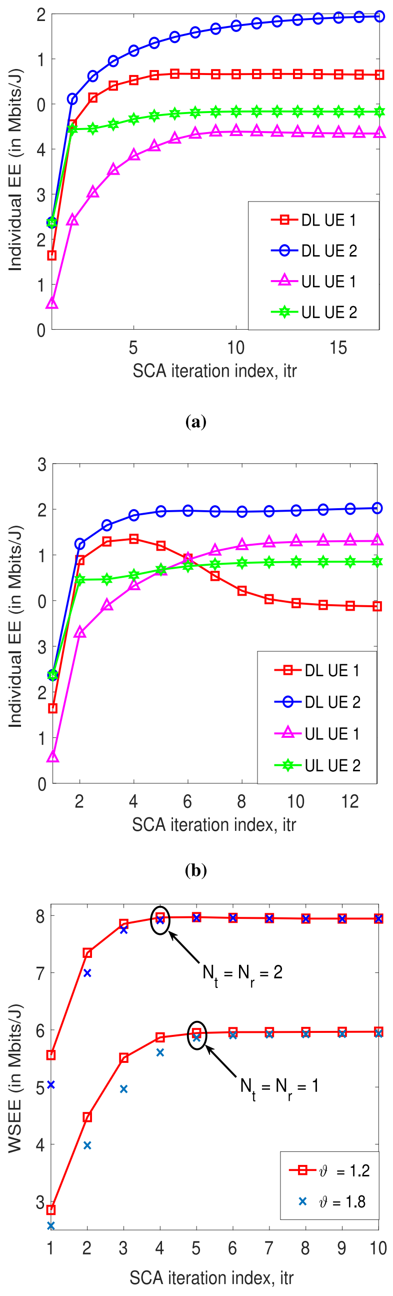
quantization bits ν. We consider $M = 32$ APs, $K_d = 12$ downlink UEs , $K_u = 8$ uplink UEs, and $p_d = 2p_u = 30$ dBm power for downlink and uplink, $N_t = N_r = \{8, 16\}$ transmit and receive antennas on each AP, and fronthaul capacities $C_{\text{fh}} = \{10, 100\}$ Mbps. We observe that for both antenna configurations, sum SE increases with increase in ν initially and then saturates. Increasing ν reduces the quantization distortion and attenuation, which improves the sum SE. This effect, however, saturates as after a limit most of the information is retrieved. We observe that reducing the fronthaul capacity from $C_{\text{fh}} = 100$ Mbps to $C_{\text{fh}} = 10$ Mbps reduces the sum SE slightly, as the procedure outlined in Section II-D *fairly* retains the AP-UE links with the highest channel gains and helps maintain the sum SE.

Sum SE - impact of channel estimation error: We know that the channel estimation error is a function of pilot transmit power p_t. We now vary p_t and evaluate its impact on the sum SE for a full-duplex cell-free massive MIMO system in Fig. 3(c). For this study, we considered $M = 32$ APs, $K_d = 12$ downlink UEs, $K_u = 8$ uplink UEs and transmit power $p_d = 2p_u = 30$ dBm. We see that the sum SE increases for $p_t \leq -10$ dB but saturates beyond that. This is because the channel estimation error reduces with increase in pilot power till $p_t = -10$ dB. Any further increase in p_t, only marginally reduces the channel estimation error, which does not affect the sum SE. Our choice of $p_t = 0.2$ W in the numerical studies is, therefore, practical.

WSEE metric - influence of weights: We now demonstrate that the WSEE metric can accommodate the heterogeneous EE requirements of both uplink and downlink UEs. For this study, we consider a particular realization of a FD CF mMIMO system with a transmit power $p_d = 2p_u = 30$ dBm, $M = 32$ APs, $K_d = K_u = K/2 = 2$ uplink and downlink UEs and $N_t = N_r = N = 2$ transmit and receive antennas on each AP, with QoS constraints $S_{ok} = S_{ol} = 0.1$ bits/s/Hz. We plot the individual EEs of the uplink (UL) and downlink (DL) UEs versus the SCA iteration index for centralized WSEE maximization, using Algorithm 3, for two different combinations of UE weights. Weights w_1 and w_2 are associated with DL UE 1 and DL UE 2, while weights w_3 and w_4 are associated with UL UE 1 and UL UE 2, respectively.

We plot in Fig. 4(a) and Fig. 4(b) the individual EEs of UL and DL UEs, with: i) equal weights ($w_1 = w_2 = w_3 = w_4 = 0.25$), and ii) $w_1 = 0.08$, $w_2 = 0.02$, $w_3 = 0.5$, $w_4 = 0.4$, respectively. In Fig. 4(a), with equal weights, UEs attain an EE depending on their relative channel conditions, which clearly indicates that in terms of channel conditions, DL UE 2 $\gg$ DL UE 1 $>$ UL UE 2 $>$ UL UE 1. In Fig. 4(b), the weights are chosen in an order which is opposite to the channel conditions. The EEs of the UL UEs now dominate the EE of DL UE 1, while reversing their relative order. The DL UE 2, with excellent channel, still attains a high EE, although lower than in Fig. 4(a).

Convergence of decentralized ADMM algorithm: We plot in Fig. 4(c) the WSEE obtained using decentralized Algorithm 4 with SCA iteration index. We consider $M = 10$ APs, $K_u = K_d = K/2 = 2$ uplink and downlink UEs and $N_t = N_r = \{1, 2\}$ transmit and receive antennas on each AP at transmit power $p_d = 2p_u = p = 30$ dBm. We assume the following: i) penalty parameters $\rho_C = \rho_\theta = 0.1$; ii) penalty parameter update threshold factor $\mu = 10$; iii) ADMM convergence threshold $\epsilon_{\text{ADMM}} = 0.01$; and iv) SCA convergence threshold $\epsilon_{\text{SCA}} = 0.001$. We consider two values of the penalty update parameter: $\vartheta = \{1.2, 1.8\}$. We note that the algorithm in both cases converges marginally quicker with $\vartheta = 1.2$. A smaller penalty update parameter is therefore



(a)

(b)

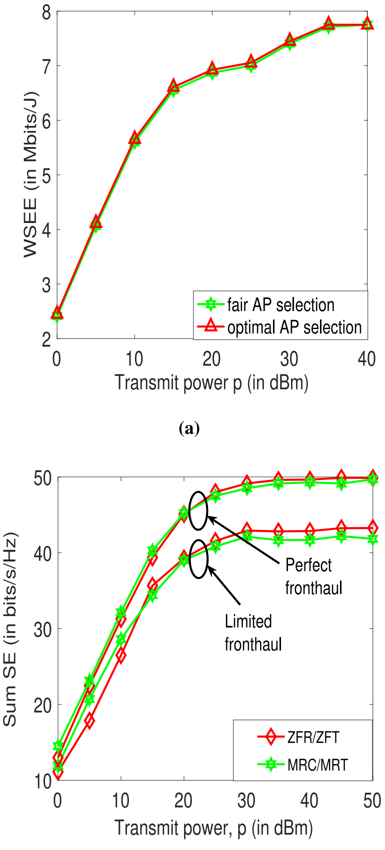
(c)

FIGURE 4. **Effect of UE priorities on individual EEs with $M = 32$, $K_d = K_u = 2$, $N_t = N_r = 2$ and $S_{ok} = S_{ol} = 0.1$ bits/s/Hz: (a) $w_1 = w_2 = w_3 = w_4 = 0.25$, (b) $w_1 = 0.08$, $w_2 = 0.02$, $w_3 = 0.5$, $w_4 = 0.4$; c) Convergence of decentralized algorithm.**

beneficial as then changes in the penalty parameters are not too abrupt, and a bad ADMM iteration which causes the primal and dual residues to diverge is, consequently, not overly responded to [37]. We therefore fix $\vartheta = 1.2$ for the rest of the simulations.

Comparison with existing schemes: We now compare our proposed FD CF mMIMO WSEE optimization strategy with some existing approaches. In particular, we compare the

- proposed fair AP selection algorithm, Algorithm 1, with the optimal AP selection scheme proposed in [29].



(a)

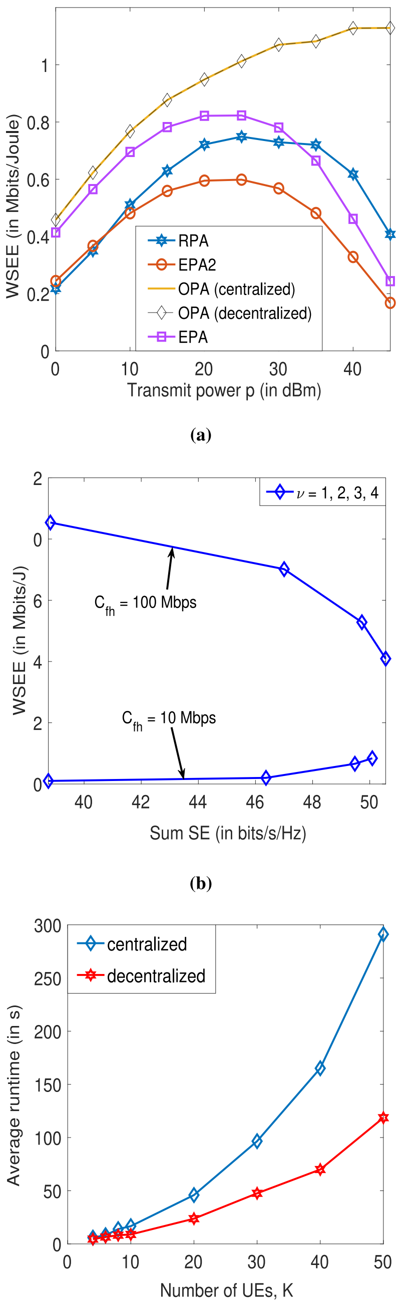
(b)

FIGURE 5. **(a) WSEE comparison between fair and optimal AP selection algorithms, (b) SE comparison between MRC/MRT and ZFR/ZFT transceivers, vs maximum transmit power p with $M = 32$ APs, $N_t = N_r = 8$ transmit and receive antennas, $K_d = 12$ downlink users and $K_u = 8$ uplink users.**

- maximum-ratio combining (MRC)/maximal ratio transmission (MRT) considered herein with zero-forcing reception (ZFR)/ zero-forcing transmission (ZFT) [39].

We observe from Fig. 5(a) that the proposed fair AP selection approach has almost as well as the optimal one in [29]. The proposed procedure efficiently eliminates the AP-UE links that do not have sufficient channel gain and thus contribute little to the system throughput while consuming a significant amount of power. Turning off APs according to the optimal AP selection procedure in [29], thus only provides marginally better WSEE.

MRC/MRT and ZFR/ZFT comparison: For this study, we considered a FD CF mMIMO system with $M = 32$ multi-antenna APs having $N_t = N_r = 8$ transmit and receive antennas each, $K_d = 12$ downlink UEs and $K_u = 8$ uplink UEs. We consider two fronthaul cases: i) perfect high-capacity with $\tilde{a} = \tilde{b} = 1$, and ii) limited $C_{\text{fh}} = 10$ Mbps capacity with $\nu = 2$ quantization bits. We see from Fig. 5(b) that for both fronthaul capacities, the MRC/MRT transceiver for the scenario considered herein, although slightly inferior at high transmit power, performs reasonably well when compared with computationally-intensive ZFR/ZFT transceiver.



(a)

(b)

(c)

FIGURE 6. **WSEE vs (a) Maximum transmit power and (b) Sum SE by varying $\nu = 1$ to 4, with $M = 32$, $K_d = K_u = 10$, $N_t = N_r = 2$ and $S_{ok} = S_{ol} = 0.1$ bits/s/Hz; c) Comparison of per-iteration runtime for decentralized and centralized algorithms.**

WSEE variation with parameters: We now vary WSEE with important system parameters and obtain crucial insights into energy-efficient FD CF mMIMO system designing. We consider $M = 32$ APs, $N_t = N_r = N = 8$ AP transmit and receive antennas, $K_d = 12$ downlink UEs, $K_u = 8$ uplink UEs and QoS constraints $S_{ok} = S_{ol} = 0.1$ bits/s/Hz, unless mentioned otherwise.

We plot in Fig. 6(a) the WSEE by simultaneously varying downlink and uplink transmit power as $p_d = 2p_u = p$. We consider centralized and decentralized optimal power allocation (OPA) approaches from Algorithm 3 and Algorithm 4, respectively. We compare them with three sub-optimal power allocation schemes: i) equal power allocation of type 1, labeled as "EPA 1", where $\eta_{mk} = (bN_t(\sum_{k\in\kappa_{dm}} \gamma^d_{mk}))^{-1}, \forall k \in \kappa_{dm}$ and $\theta_l = 1$ [18], [19], ii) equal power allocation of type 2, labeled as "EPA 2", where $\eta_{mk} = (bN_t K_{dm}\gamma^d_{mk})^{-1}, \forall k \in \kappa_{dm}$ and $\theta_l = 1$ [18], and iii) random power allocation, labeled as "RPA", where power control coefficients are chosen randomly from a uniform distribution between 0 and the "EPA 1" value. We note that the existing literature has not yet optimized the WSEE metric for CF mMIMO systems, and hence we can only compare with above sub-optimal schemes. Further, the decentralized ADMM approach, with lower computational complexity, has the same WSEE as that of the centralized one. Also, both decentralized and centralized approaches far outperform the baseline schemes.

We next characterize in Fig. 6(b) the joint variation of WSEE and sum SE with the number of quantization bits ν in the fronthaul links. The WSEE is obtained using decentralized Algorithm 4. We consider transmit power $p_d = 2p_u = p = 30$ dBm and take two different cases: i) high fronthaul capacity, $C_{\text{fh}} = 100$ Mbps, which is sufficiently high to support all the UEs, and ii) limited fronthaul capacity, $C_{\text{fh}} = 10$ Mbps, which limits the number of UEs a single AP can serve. We observe that for $C_{fh} = 100$ Mbps, the WSEE falls with increase in ν, even though the corresponding sum SE increases. For $C_{fh} = 10$ Mbps, both sum SE and WSEE simultaneously increase with increase in ν. To explain this behavior, we note from Fig. 3(b) that increasing ν improves the sum SE for $C_{\text{fh}} = 100$ Mbps and $C_{\text{fh}} = 10$ Mbps. For $C_{fh} = 100$ Mbps, the APs serve all the UEs, i.e., $K_{dm} = K_d$ and $K_{um} = K_u$, so increasing ν linearly increases the fronthaul data rate, R_{fh} (see (7)). This, as seen from (18), increases the traffic-dependent fronthaul power consumption. Using lower number (1-2) of quantization bits is therefore more energy-efficient, as it provides sufficiently good SE with a low energy consumption. However, for $C_{fh} = 10$ Mbps, K_{um} and K_{dm} have an upper limit, given by (9), which is inversely related to ν. The product, $\nu(K_{um}+K_{dm})$, remains nearly constant for all values of ν. Thus, R_{fh} (see (7)) doesn't increase with increase in ν and remains close to the capacity, C_{fh}. The traffic-dependent fronthaul power consumption, given in (18), hence, remains close to P_{ft}. A higher number of quantization bits $(3-4)$ therefore provides a higher sum SE and hence, also maximizes the WSEE.

Latency: The *per-iteration* complexity of the *decentralized* Algorithm 4, as observed earlier in Section IV-C, is lower than the *centralized* Algorithm 3. We now demonstrate the same by comparing their *per-iteration* runtime. For this simulation, as shown in Fig. 6(c), we consider an FD CF mMIMO system with $M = 32$ APs, each having $N_t = N_r = 8$ transmit and receive antennas, and plot the average runtime of each iteration by varying the total number of UEs, K, with $K_d = K_u = K/2$. We note that the decentralized algorithm has significantly lower per-iteration runtime, particularly for

TABLE 2. Optimal uniform quantization parameters.

ν	$\tilde{\Delta}_{\text{opt}}$	$\mathbb{E}\{\tilde{\varsigma}_d^2\} = \tilde{b} - \tilde{a}^2$	$\tilde{a}$
1	1.596	0.2313	0.6366
2	0.9957	0.10472	0.88115
3	0.586	0.036037	0.96256
4	0.3352	0.011409	0.98845
5	0.1881	0.003482	0.996505
6	0.1041	0.0010389	0.99896

large K. Both these algorithms require only large-scale channel coefficients and hence need to be executed only once in *hundreds* of coherence intervals.

VI. CONCLUSION

We derived a SE lower bound for a FD CF mMIMO wireless system with optimal uniform fronthaul quantization. Using a *two-layered* approach, we optimized WSEE using SCA framework which in each iteration solves a GCP either centrally or decentrally using ADMM. We showed how WSEE incorporates EE requirements of different UEs. We analytically and numerically demonstrated the convergence of decentralized algorithm. We showed that it achieves the same WSEE as the centralized approach with a much reduced computational complexity.

APPENDIX A

We use the optimal uniform quantization model from [15], [19]. Using Bussgang decomposition [40], the quantization function $\mathcal{Q}(x) = \tilde{a}x + \sqrt{p_x}\tilde{\varsigma}_d$, where $p_x = \mathbb{E}\{|x|^2\}$ is the power of the unquantized signal x, $\tilde{a} = \frac{1}{p_x}\int_{\mathcal{X}} xh(x)f_X(x)dx$, $\tilde{b} = \frac{1}{p_x}\int_{\mathcal{X}} h^2(x)f_X(x)dx$ and $\tilde{\varsigma}_d$ is the normalized distortion whose power is given as $\mathbb{E}\{\tilde{\varsigma}_d^2\} = \tilde{b} - \tilde{a}^2$. Here $h(x)$ is the mid-rise uniform quantizer with $L = 2^\nu$ quantization levels rising in steps of size $\tilde{\Delta}$, and ν being the number of quantization bits. The signal-to-distortion ratio $\text{SDR} = \frac{\mathbb{E}\{(\tilde{a}x)^2\}}{p_x\mathbb{E}\{\tilde{\varsigma}_d^2)\}} = \frac{\tilde{a}^2}{\tilde{b}-\tilde{a}^2}$. The optimal step-size $\tilde{\Delta}_{\text{opt}}$ maximizes the SDR for a given ν. The optimal $\tilde{a}$ and $\tilde{b}$ values are calculated using the optimal $\tilde{\Delta}_{\text{opt}}$ for each value of ν, and are given in Table 2 [15].

APPENDIX B

We now derive the achievable SE expression for the kth downlink UE in (14). From Section II-B, we know that $\mathbf{g}_{mk}^d = \hat{\mathbf{g}}_{mk}^d + \mathbf{e}_{mk}^d$, where $\hat{\mathbf{g}}_{mk}^d$ and $\mathbf{e}_{mk}^d$ are independent and $\mathbb{E}\{\|\hat{\mathbf{g}}_{mk}^d\|^2\} = N_t\gamma_{mk}^d$. We can express the desired signal for the kth downlink UE as

$$\begin{aligned}
&\mathbb{E}\left\{|\text{DS}_k^d|^2\right\} \\
&= \tilde{a}^2\rho_d\mathbb{E}\left\{|\sum_{m\in\mathcal{M}_k^d}\sqrt{\eta_{mk}}\mathbb{E}\left\{\left(\mathbf{g}_{mk}^d\right)^T\left(\hat{\mathbf{g}}_{mk}^d\right)^*\right\}s_k^d|^2\right\} \\
&= \tilde{a}^2N_t^2\rho_d\left(\sum_{m\in\mathcal{M}_k^d}\sqrt{\eta_{mk}}\gamma_{mk}^d\right)^2. \qquad (39)
\end{aligned}$$

We now calculate the beamforming uncertainty for the kth downlink UE as follows

$$\begin{aligned}
&\mathbb{E}\left\{\left|\text{BU}_k^d\right|^2\right\} \\
&= \tilde{a}^2\rho_d\sum_{m\in\mathcal{M}_k^d}\eta_{mk}\mathbb{E}\left\{\left|\left(\mathbf{g}_{mk}^d\right)^T\left(\hat{\mathbf{g}}_{mk}^d\right)^* - \mathbb{E}\left\{\left(\mathbf{g}_{mk}^d\right)^T\left(\hat{\mathbf{g}}_{mk}^d\right)^*\right)\right\}\right|^2\right\} \\
&\overset{(a)}{=} \tilde{a}^2\rho_d\sum_{m\in\mathcal{M}_k^d}\eta_{mk}\Bigg(N_t(N_t+1)\left(\gamma_{mk}^d\right)^2 \\
&\quad + N_t\gamma_{mk}^d\left(\beta_{mk}^d - \gamma_{mk}^d\right) - N_t^2\left(\gamma_{mk}^d\right)^2\Bigg) \\
&= \tilde{a}^2N_t\rho_d\sum_{m\in\mathcal{M}_k^d}\eta_{mk}\beta_{mk}^d\gamma_{mk}^d. \qquad (40)
\end{aligned}$$

Equality (a) is because i) $\hat{\mathbf{g}}_{mk}^d$ are zero-mean and uncorrelated; and ii) $\mathbb{E}\{\|\hat{\mathbf{g}}_{mk}^d\|^4\} = N_t(N_t+1)(\gamma_{mk}^d)^2$ [12] and $\mathbb{E}\{\|\mathbf{e}_{mk}^d\|^2\} = (\beta_{mk}^d - \gamma_{mk}^d)$.

We now simplify MUI for the kth downlink UE:

$$\begin{aligned}
&\mathbb{E}\left\{\left|\text{MUI}_k^d\right|^2\right\} \\
&= \tilde{a}^2\rho_d\sum_{m=1}^{M}\sum_{q\in\mathcal{K}_{dm}\backslash k}\eta_{mq}\mathbb{E}\left\{\left|\left(\mathbf{g}_{mk}^d\right)^T\left(\hat{\mathbf{g}}_{mq}^d\right)^*\right|^2\right\} \\
&\overset{(a)}{=} \tilde{a}^2N_t\rho_d\sum_{m=1}^{M}\sum_{q\in\mathcal{K}_{dm}\backslash k}\beta_{mk}^d\eta_{mq}\gamma_{mq}^d. \qquad (41)
\end{aligned}$$

Equality (a) is because: i) $\hat{\mathbf{g}}_{mq}^d$ and $\mathbf{g}_{mk}^d$ are mutually independent; and ii) $\mathbb{E}\{|(\mathbf{g}_{mk}^d)^T(\hat{\mathbf{g}}_{mq}^d)^*|^2\} = \mathbb{E}\{(\hat{\mathbf{g}}_{mq}^d)^T\mathbb{E}\{(\mathbf{g}_{mk}^d)^*(\mathbf{g}_{mk}^d)^T\}(\hat{\mathbf{g}}_{mq}^d)^*\} = N_t\beta_{mk}^d\gamma_{mq}^d$.

We next calculate UDI for the kth downlink UE:

$$\mathbb{E}\left\{\left|\text{UDI}_k^d\right|^2\right\} = \rho_u\sum_{l=1}^{K_u}\mathbb{E}\left\{|h_{kl}|^2\right\}\theta_l = \rho_u\sum_{l=1}^{K_u}\tilde{\beta}_{kl}\theta_l. \qquad (42)$$

We express the total quantization distortion (TQD) for the kth downlink UE as follows

$$\begin{aligned}
\mathbb{E}\left\{\left|\text{TQD}_k^d\right|^2\right\} &\approx \rho_d\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}}\mathbb{E}\left\{\left|\left(\mathbf{g}_{mk}^d\right)^T\left(\hat{\mathbf{g}}_{mq}^d\right)^*\varsigma_{mq}^d\right|^2\right\} \\
&\overset{(a)}{=} \left(\tilde{b} - \tilde{a}^2\right)N_t\rho_d\sum_{m=1}^{M}\sum_{q\in\kappa_{dm}}\beta_{mk}^d\eta_{mq}\gamma_{mq}^d. \qquad (43)
\end{aligned}$$

Equality (a) is because: i) $\mathbb{E}\{|\varsigma_{mk}^d|^2\} = (\tilde{b} - \tilde{a}^2)\eta_{mk}$; ii) distortion ς_{mq}^d is independent of channels $\mathbf{g}_{mk}^d$ and $\hat{\mathbf{g}}_{mq}^d$; and iii) $\mathbb{E}\{|(\mathbf{g}_{mk}^d)^T(\hat{\mathbf{g}}_{mq}^d)^*\varsigma_{mq}^d|^2\} = (\tilde{b} - \tilde{a}^2)\eta_{mq}\beta_{mk}^d\mathbb{E}\{(\hat{\mathbf{g}}_{mq}^d)^T(\hat{\mathbf{g}}_{mq}^d)^*\} = (\tilde{b} - \tilde{a}^2)N_t\beta_{mk}^d\eta_{mq}\gamma_{mq}^d$. The result in (14) follows from the expression for the achievable SE lower bound

$$S_k^d = \tau_f\log_2\left(1 + \frac{\mathbb{E}\left\{\left|\text{DS}_k^d\right|^2\right\}}{\left\{\begin{array}{r}\mathbb{E}\left\{\left|\text{BU}_k^d\right|^2\right\} + \mathbb{E}\left\{\left|\text{MUI}_k^d\right|^2 + \mathbb{E}\{|\text{UDI}_k^d|^2\}\right\} \\ + \mathbb{E}\left\{\left|\text{TQD}_k^d\right|^2\right\} + \mathbb{E}\left\{\left|w_k^d\right|^2\right\}\end{array}\right\}}\right).$$

We now derive the achievable SE expression for the lth uplink UE in (15). We know from Section II-B that $\boldsymbol{g}_{ml}^u = \hat{\boldsymbol{g}}_{ml}^u + \boldsymbol{e}_{ml}^u$, where $\hat{\boldsymbol{g}}_{ml}^u$ and $\boldsymbol{e}_{ml}^u$ are independent and $\mathbb{E}\{\|\hat{\boldsymbol{g}}_{ml}^u\|^2\} = N_r\gamma_{ml}^u$. We can express the desired signal for the lth uplink UE as given next

$$\mathbb{E}\Big\{\big|\mathrm{DS}_l^u\big|^2\Big\} = \mathbb{E}\left\{\left|\tilde{a}\sum_{m\in\mathcal{M}_l^u}\sqrt{\rho_u}\mathbb{E}\Big\{\sqrt{\theta_l}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\big(\hat{\boldsymbol{g}}_{ml}^u + \boldsymbol{e}_{ml}^u\big)s_l^u\Big\}\right|^2\right\}$$
$$= \tilde{a}^2N_r^2\rho_u\theta_l\left(\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u\right)^2. \qquad (44)$$

The beamforming uncertainty for the lth uplink UE is

$$\mathbb{E}\Big\{\big|\mathrm{BU}_l^u\big|^2\Big\}$$
$$= \tilde{a}^2\rho_u\theta_l\sum_{m\in\mathcal{M}_l^u}\mathbb{E}\left\{\left\|\left(\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{ml}^u - \mathbb{E}\Big\{\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{ml}^u\Big\}\right)\right\|^2\right\}$$
$$\overset{(a)}{=} \tilde{a}^2\rho_u\theta_l\sum_{m\in\mathcal{M}_l^u}\left(\mathbb{E}\Big\{\|\hat{\boldsymbol{g}}_{ml}^u\|^4\Big\} + \mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{e}_{ml}^u\right|^2\right\} - N_r^2\big(\gamma_{ml}^u\big)^2\right)$$
$$\overset{(b)}{=} \tilde{a}^2\rho_uN_r\theta_l\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u\beta_{ml}^u. \qquad (45)$$

Equality (a) is because: i) $\boldsymbol{e}_{ml}^u$ and $\hat{\boldsymbol{g}}_{ml}^u$ are zero-mean and uncorrelated; ii) $\mathbb{E}\{|\hat{\boldsymbol{g}}_{ml}^u|^2\} = N_r\gamma_{ml}^u$. Equality ($b$) is because $\mathbb{E}\{\|\hat{\boldsymbol{g}}_{ml}^u\|^4\} = N_r(N_r+1)(\gamma_{ml}^u)^2$ [12] and $\mathbb{E}\{\|\boldsymbol{e}_{ml}^u\|^2\} = (\beta_{ml}^u - \gamma_{ml}^u)$.

We simplify the MUI for the lth uplink UE as

$$\mathbb{E}\Big\{\big|\mathrm{MUI}_l^u\big|^2\Big\} = \tilde{a}^2\rho_u\sum_{m\in\mathcal{M}_l^u}\sum_{q=1,q\neq l}^{K_u}\theta_q\mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{mq}^u\right|^2\right\}$$
$$\overset{(a)}{=} \tilde{a}^2\rho_uN_r\sum_{m\in\mathcal{M}_l^u}\sum_{q=1,q\neq l}^{K_u}\gamma_{ml}^u\beta_{mq}^u\theta_q. \qquad (46)$$

Equality (a) is obtained by using these facts: i) $\hat{\boldsymbol{g}}_{ml}^u$, $\boldsymbol{g}_{mq}^u$ are mutually independent; and ii)

$$\mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{mq}^u\right|^2\right\} = \mathbb{E}\left\{\Big(\boldsymbol{g}_{mq}^u\Big)^H\mathbb{E}\Big\{\big(\hat{\boldsymbol{g}}_{ml}^u\big)\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\Big\}\boldsymbol{g}_{mq}^u\right\}$$
$$= \gamma_{ml}^u\mathbb{E}\Big\{||\boldsymbol{g}_{mq}^u||^2\Big\} = N_r\gamma_{ml}^u\beta_{mq}^u. \quad (47)$$

We next obtain the noise power for the lth uplink UE as

$$\mathbb{E}\Big\{\big|\mathrm{N}_l^u\big|^2\Big\} = \tilde{a}^2\sum_{m\in\mathcal{M}_l^u}\mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{w}_m^u\right|^2\right\} = \tilde{a}^2N_r\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u, \text{ where}$$
$$\mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{w}_m^u\right|^2\right\} = \mathbb{E}\Big\{\big(\boldsymbol{w}_m^u\big)^H\mathbb{E}\Big\{\hat{\boldsymbol{g}}_{ml}^u\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\Big\}\boldsymbol{w}_m^u\Big\} = N_r\gamma_{ml}^u. \qquad (48)$$

The undistorted MR-combined uplink signal at the mth AP is expressed as

$$\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{y}_m^u$$
$$= \sum_{q=1}^{K_u}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{mq}^ux_q^u + \sum_{i=1}^{M}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{H}_{mi}\boldsymbol{x}_i^d + \big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{w}_m^u$$
$$= \underbrace{\sqrt{\rho_u}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{ml}^u\sqrt{\theta_l}s_l^u}_{\text{message signal}} + \underbrace{\sqrt{\rho_u}\sum_{q=1,q\neq l}^{K_u}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{g}_{mq}^u\sqrt{\theta_q}s_q^u}_{\text{multi-user interference, MUI}_l^u}$$
$$+ \underbrace{\sqrt{\rho_d}\sum_{i=1}^{M}\sum_{k\in\kappa_{di}}\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{H}_{mi}\Big(\hat{\boldsymbol{g}}_{ik}^d\Big)^*\Big(\tilde{a}\sqrt{\eta_{ik}}s_k^d + \varsigma_{ik}^d\Big)}_{\text{intra-/inter-AP residual interference, RI}_l^u}$$
$$+ \underbrace{\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{w}_m^u}_{\text{additive noise at APs, N}_l^u}. \qquad (49)$$

We assume, similar to [19], that the quantization distortion is uncorrelated across the fronthaul links. The TQD power for the lth uplink UE is accordingly expressed as

$$\mathbb{E}\Big\{\big|\mathrm{TQD}_l^u\big|^2\Big\} \approx \sum_{m\in\mathcal{M}_l^u}\mathbb{E}\Big\{|\zeta_{ml}^u|^2\Big\}$$
$$\approx \Big(\tilde{b} - \tilde{a}^2\Big)\sum_{m\in\mathcal{M}_l^u}\mathbb{E}\left\{\left|\big(\hat{\boldsymbol{g}}_{ml}^u\big)^H\boldsymbol{y}_m\right|^2\right\}.$$

Using arguments similar to (44)-(48), the contributions of the message signal (DS + BU), MUI and noise (N) to the TQD for the lth uplink UE are

$$\mathbb{E}\Big\{\big|\mathrm{TQD}_l^u\big|^2\Big\}_{\text{DS+BU}}$$
$$\approx \Big(\tilde{b} - \tilde{a}^2\Big)N_r\rho_u\theta_l\left(N_r\sum_{m\in\mathcal{M}_l^u}\big(\gamma_{ml}^u\big)^2 + \sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u\beta_{ml}^u\right).$$
$$\mathbb{E}\Big\{\big|\mathrm{TQD}_l^u\big|^2\Big\}_{\text{MUI}} \approx \Big(\tilde{b} - \tilde{a}^2\Big)N_r\rho_u\sum_{m\in\mathcal{M}_l^u}\sum_{q=1,q\neq l}^{K_u}\gamma_{ml}^u\beta_{mq}^u\theta_q,$$
$$\mathbb{E}\Big\{\big|\mathrm{TQD}_l^u\big|^2\Big\}_{\text{N}} \approx \Big(\tilde{b} - \tilde{a}^2\Big)N_r\sum_{m\in\mathcal{M}_l^u}\gamma_{ml}^u.$$

To accurately model the RI with limited fronthaul capacity and compute the corresponding power, as well as its contribution to the quantization distortion, we propose a lemma.

Lemma 2: The intra-/inter-AP RI power and the RI contribution to the TQD power for the lth uplink UE in a FD CF mMIMO system with MRT/MRC transceiver are expressed as

$$\mathbb{E}\Big\{\big|\mathrm{RI}_l^u\big|^2\Big\}$$
$$= \tilde{a}^2\tilde{b}N_rN_t\rho_d\sum_{i=1}^{M}\sum_{k\in\kappa_{di}}\gamma_{ml}^u\gamma_{ik}^d\beta_{\text{RI},mi}\gamma_{\text{RI}}\eta_{ik}N_r\gamma_{ml}^u. \qquad (50)$$
$$\mathbb{E}\Big\{\big|\mathrm{TQD}_l^u\big|^2\Big\}_{\text{RI}}$$
$$\approx \Big(\tilde{b} - \tilde{a}^2\Big)\tilde{b}N_rN_t\rho_d\sum_{m\in\mathcal{M}_l^u}\sum_{i=1}^{M}\sum_{k\in\kappa_{di}}\gamma_{ml}^u\beta_{\text{RI},mi}\gamma_{\text{RI}}\eta_{ik}\gamma_{ik}^d. \quad (51)$$

Proof: We express the RI power of the undistorted, MR combined received signal for the lth uplink UE as

$$\begin{aligned}
&\mathbb{E}\left\{\left|\widetilde{\mathrm{RI}}_l^u\right|^2\right\} \\
&= \rho_d \sum_{i=1}^{M} \sum_{k\in\kappa_{di}} \mathbb{E}\left\{\left|\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\left(\tilde{a}\sqrt{\eta_{ik}}s_k^d + \zeta_{ik}^d\right)\right|^2\right\} \\
&\overset{(a)}{=} \rho_d \sum_{i=1}^{M} \sum_{k\in\kappa_{di}} \mathbb{E}\left\{\left|\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\right|^2 \tilde{b}\eta_{ik}\right\} \\
&\overset{(b)}{=} \tilde{b}N_rN_t\rho_d \sum_{i=1}^{M} \sum_{k\in\kappa_{di}} \gamma_{ml}^u\gamma_{ik}^d\beta_{\mathrm{RI},mi}\gamma_{\mathrm{RI}}\eta_{ik}N_r\gamma_{ml}^u.
\end{aligned}$$

Equality (a) is because signal $\tilde{a}\sqrt{\eta_{ik}}s_k^d$ and quantization noise ς_{ik}^d, are uncorrelated, and $\mathbb{E}\{|\varsigma_{ik}^d|^2\} = (\tilde{b} - \tilde{a}^2)\eta_{ik}$. Equality (b) is because: i) $\hat{\boldsymbol{g}}_{ml}^u$, $\boldsymbol{H}_{mi}$ and $\hat{\boldsymbol{g}}_{mk}^d$ are mutually independent,

$$\begin{aligned}
\text{ii) } &\mathbb{E}\left\{\left|\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H \boldsymbol{H}_{mi}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\right|^2\right\} \\
&= \mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{ik}^d\right)^T \mathbb{E}\left\{\boldsymbol{H}_{mi}^H \mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{ml}^u\right)\left(\hat{\boldsymbol{g}}_{ml}^u\right)^H\right\}\boldsymbol{H}_{mi}\right\}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\right\} \\
&= \gamma_{ml}^u \mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{ik}^d\right)^T \mathbb{E}\left\{\boldsymbol{H}_{mi}^H\boldsymbol{H}_{mi}\right\}\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\right\} \\
&= N_r\gamma_{ml}^u\beta_{\mathrm{RI},mi}\gamma_{\mathrm{RI}}\mathbb{E}\left\{\left(\hat{\boldsymbol{g}}_{ik}^d\right)^T\left(\hat{\boldsymbol{g}}_{ik}^d\right)^*\right\} \\
&= N_rN_t\gamma_{ml}^u\gamma_k^d\beta_{\mathrm{RI},mi}\gamma_{\mathrm{RI}}. \qquad (52)
\end{aligned}$$

We obtain the i) attenuated intra-/inter-AP RI power as $\mathbb{E}\{|\mathrm{RI}_l^u|^2\} = \tilde{a}^2\mathbb{E}\{|\widetilde{\mathrm{RI}}_l^u|^2\}$; and intra-/inter-AP RI contribution to the TQD power as $\mathbb{E}\{|\mathrm{TQD}_l^u|^2\}_{\mathrm{RI}} \approx (\tilde{b} - \tilde{a}^2)\sum_{m\in\mathcal{M}_l^u}\mathbb{E}\{|\widetilde{\mathrm{RI}}_l^u|^2\}$. ∎

The total quantization distortion for the lth uplink UE is given as $\mathbb{E}\{|\mathrm{TQD}_l^u|^2\} = \mathbb{E}\{|\mathrm{TQD}_l^u|^2\}_{\mathrm{DS+BU}} + \mathbb{E}\{|\mathrm{TQD}_l^u|^2\}_{\mathrm{MUI}} + \mathbb{E}\{|\mathrm{TQD}_l^u|^2\}_{\mathrm{RI}} + \mathbb{E}\{|\mathrm{TQD}_l^u|^2\}_{\mathrm{N}}$.

The result in (15) follows from the expression

$$S_l^u = \tau_f \log_2\left(1 + \frac{\mathbb{E}\left\{\left|\mathrm{DS}_l^u\right|^2\right\}}{\left\{\begin{aligned}&\mathbb{E}\left\{\left|\mathrm{BU}_l^u\right|^2\right\} + \mathbb{E}\left\{\left|\mathrm{MUI}_l^u\right|^2\right\} + \mathbb{E}\left\{\left|\mathrm{RI}_l^u\right|^2\right\} \\ &\qquad + \mathbb{E}\left\{\left|\mathrm{TQD}_l^u\right|^2\right\} + \mathbb{E}\left\{\left|\mathrm{N}_l^u\right|^2\right\}\end{aligned}\right\}}\right).$$

REFERENCES

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, Nov. 2010.

[2] Ö. T. Demir, E. Björnson, and L. Sanguinetti, "Foundations of user-centric cell-free massive MIMO," *Found. Trends®Signal Process.*, vol. 14, nos. 3–4, pp. 162–472, 2020. [Online]. Available: http://dx.doi.org/10.1561/2000000109

[3] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[4] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and H. Yang, "Cell-free massive MIMO systems," in *Proc. 49th Asilomar Conf. Signals Syst. Comput.*, 2015, pp. 695–699.

[5] R. Chopra, C. R. Murthy, and A. K. Papazafeiropoulos, "Uplink performance analysis of cell-free mMIMO systems under channel aging," *IEEE Commun. Lett.*, vol. 25, no. 7, pp. 2206–2210, Jul. 2021.

[6] T. Riihonen, S. Werner, and R. Wichman, "Mitigation of loopback self-interference in full-duplex MIMO relays," *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5983–5993, Dec. 2011.

[7] X. Xia, D. Zhang, K. Xu, W. Ma, and Y. Xu, "Hardware impairments aware transceiver for full-duplex massive MIMO relaying," *IEEE Trans. Signal Process.*, vol. 63, no. 24, pp. 6565–6580, Dec. 2015.

[8] M. Jain *et al.*, "Practical, real-time, full duplex wireless," in *Proc. 17th Annu. Int. Conf. Mobile Comput. Netw. (MobiCom)*, 2011, pp. 301–312.

[9] D. Bharadia, E. McMillin, and S. Katti, "Full duplex radios," *ACM Sigcomm Comput. Commun. Rev.*, vol. 43, no. 4, pp. 375–386, 2013.

[10] Y. Jang, K. Min, S. Park, and S. Choi, "Spatial resource utilization to maximize uplink spectral efficiency in full-duplex massive MIMO," in *Proc. IEEE Int. Conf. Commun.(ICC)*, 2015, pp. 1583–1588.

[11] Y. Li, P. Fan, A. Leukhin, and L. Liu, "On the spectral and energy efficiency of full-duplex small-cell wireless systems with massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 66, no. 3, pp. 2339–2353, Mar. 2017.

[12] 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)*, 2019, pp. 1–6.

[13] D. Wang, M. Wang, P. Zhu, J. Li, J. Wang, and X. You, "Performance of network-assisted full-duplex for cell-free massive MIMO," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1464–1478, Mar. 2020.

[14] H. V. Nguyen *et al.*, "A novel heap-based pilot assignment for full duplex cell-free massive MIMO with zero-forcing," in *Proc. IEEE Int. Conf. Commun.(ICC)*, 2020, pp. 1–6.

[15] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, M. Debbah, and P. Xiao, "Max–Min rate of cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Commun.*, vol. 67, no. 10, pp. 6796–6815, Oct. 2019.

[16] G. Femenias and F. Riera-Palou, "Cell-free millimeter-wave massive MIMO systems with limited fronthaul capacity," *IEEE Access*, vol. 7, pp. 44596–44612, Apr. 2019.

[17] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1038–1053, Feb. 2020.

[18] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[19] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 971–987, Dec. 2019.

[20] M. Alonzo, S. Buzzi, A. Zappone, and C. D'Elia, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 3, pp. 651–663, Sep. 2019.

[21] H. V. Nguyen *et al.*, "On the spectral and energy efficiencies of full-duplex cell-free massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1698–1718, Aug. 2020.

[22] A. Zappone and E. Jorswieck, *Energy Efficiency in Wireless Networks via Fractional Programming Theory*, vol. 11. Hanover, MA, USA: Now Publ., Jun. 2015. [Online]. Available: https://doi.org/10.1561/0100000088

[23] C. N. Efrem and A. D. Panagopoulos, "A framework for weighted-sum energy efficiency maximization in wireless networks," *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 153–156, Feb. 2019.

[24] E. Sharma, D. N. Amudala, and R. Budhiraja, "Energy efficiency optimization of massive MIMO FD relay with quadratic transform," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1429–1448, Feb. 2020.

[25] C. Jeon, K. Li, J. R. Cavallaro, and C. Studer, "Decentralized equalization with feedforward architectures for massive MU-MIMO," *IEEE Trans. Signal Process.*, vol. 67, no. 17, pp. 4418–4432, Sep. 2019.

[26] J. Rodríguez Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 687–700, 2020.

[27] Z. Zhou, J. Feng, Z. Chang, and X. Shen, "Energy-efficient edge computing service provisioning for vehicular networks: A consensus ADMM approach," *IEEE Trans. Veh. Technol.*, vol. 68, no. 5, pp. 5087–5099, May 2019.

[28] A. Goldsmith, *Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[29] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and load balancing optimization for energy-efficient cell-free massive MIMO networks," *IEEE Trans. Wireless Commun.*, vol. 19, no. 10, pp. 6798–6812, Oct. 2020.

[30] E. Everett, A. Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680–694, Feb. 2014.

[31] H. Q. Ngo, H. A. Suraweera, M. Matthaiou, and E. G. Larsson, "Multipair full-duplex relaying with massive arrays and linear processing," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1721–1737, Sep. 2014.

[32] Z. Zhang, Z. Ma, Z. Ding, M. Xiao, and G. K. Karagiannidis, "Full-duplex two-way and one-way relaying: Average rate, outage probability, and tradeoffs," *IEEE Trans. Wireless Commun.*, vol. 15, no. 6, pp. 3920–3933, Jun. 2016.

[33] X. Xiong, X. Wang, T. Riihonen, and X. You, "Channel estimation for full-duplex relay systems with large-scale antenna arrays," *IEEE Trans. Wireless Commun.*, vol. 15, no. 10, pp. 6925–6938, Oct. 2016.

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

[35] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends® Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2010.

[36] B. R. Marks and G. P. Wright, "Technical note–A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, 1978.

[37] B. He, H. Yang, and S. Wang, "Alternating direction method with selfadaptive penalty parameters for monotone variational inequalities," *J. Optim. Theory Appl.*, vol. 106, no. 2, p. 337–356, Aug. 2000.

[38] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*. Philadelphia, PA, USA: Soc. Ind. Appl. Math., 2001.

[39] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint power allocation and user association optimization for massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6384–6399, Sep. 2016.

[40] P. Zillmann, "Relationship between two distortion measures for memoryless nonlinear systems," *IEEE Signal Process. Lett.*, vol. 17, no. 11, pp. 917–920, Nov. 2010.

SOUMYADEEP DATTA received the B.Tech.–M.Tech. dual degree in electrical engineering from the Indian Institute of Technology Kanpur, India, in 2020. He is currently pursuing the dual Ph.D. degree with the Department of Electrical Engineering, Indian Institute of Technology Kanpur and the Department of Electrical and Computer Engineering, NYU Tandon School of Engineering, USA. His research interests include beyond-5G wireless systems, wireless networks, and cross-layer optimization.

DHEERAJ NAIDU AMUDALA (Graduate Student Member, IEEE) received the B.Tech. degree in electronics and communication engineering from JNTUACEA, Ananthapuramu, India, in 2016. He is currently pursuing the M.Tech. and Ph.D. degrees with the Department of Electrical Engineering, Indian Institute of Technology Kanpur, India.

His research interests include massive MIMO, full-duplex, wireless relaying systems, and optimization theory.

EKANT SHARMA (Member, IEEE) received the M.Tech. and Ph.D. degrees in electrical engineering from the Signal Processing, Communication and Networks Group, Department of Electrical Engineering, Indian Institute of Technology Kanpur, India, in May 2011 and May 2020, respectively. From 2011 to 2012, he was with the IBM-India Software Lab and worked as an Associate Software Engineer. From August 2019 to January 2021, he worked with the 5G Testbed Lab, Indian Institute of Technology Kanpur, where he designed base station hardware and software algorithms for 5G NR. He is currently working as an Assistant Professor with the Indian Institute of Technology Roorkee. His Ph.D. thesis received outstanding thesis award and also it was chosen for category: SPCOM Best Doctoral Dissertation—Honorable Mention at IEEE SPCOM Conference. His research interests are within the areas of wireless communications systems, with special focus on practical massive MIMO, full-duplex, relays, energy efficiency, and optimization.

ROHIT BUDHIRAJA received the M.S. degree in electrical engineering and the Ph.D. degree from the Indian Institute of Technology (IIT) Madras in 2004 and 2015, respectively. From 2004 to 2011, he worked for two startups where he designed both hardware and software algorithms, from scratch, for physical layer processing of WiMAX- and LTE-based cellular systems. He is currently an Assistant Professor with IIT Kanpur, where he is also leading an effort to design a 5G research testbed. His current research interests include the design of energy-efficient transceiver algorithms for 5G massive MIMO and full-duplex systems, robust precoder design for wireless relaying, machine learning methods for channel estimation in mm-wave systems, and spatial modulation system design. His paper was shortlisted as one of the finalists for the Best Student Paper Awards at the IEEE International Conference on Signal Processing and Communications, Bengaluru, India, in 2014. He also received the IIT Madras Research Award for the quality and quantity of research work done in the Ph.D., Early Career Research Award, and Teaching Excellence Certificate at IIT Kanpur.

SHIVENDRA S. PANWAR (Fellow, IEEE) received the Ph.D. degree in electrical and computer engineering from the University of Massachusetts, Amherst, MA, USA, in 1986. He is currently a Professor with the Electrical and Computer Engineering Department, NYU Tandon School of Engineering. He is also the Director of the New York State Center for Advanced Technology in Telecommunications, the Co-Founder of the New York City Media Lab, and a member of NYU Wireless. His research interests include the performance analysis and design of networks. His current research focuses on cross-layer research issues in wireless networks, and multimedia transport over networks. He has coauthored a textbook titled *TCP/IP Essentials: A Lab-Based Approach* (Cambridge University Press). He was a winner of the IEEE Communication Society's Leonard Abraham Prize for 2004, the ICC Best Paper Award in 2016, and the Sony Research Award. He was also co-awarded the Best Paper in 2011 Multimedia Communications Award. He has served as the Secretary for the Technical Affairs Council of the IEEE Communications Society.