User-Centric 5G Cellular Networks: Resource Allocation and Comparison With the Cell-Free Massive MIMO Approach

Stefano Buzzi, *Senior Member, IEEE*, Carmen D'Andrea, *Student Member, IEEE*, Alessio Zappone, *Senior Member, IEEE*, and Ciro D'Elia

***Abstract*— Recently, the so-called cell-free (CF) massive multiple-input multiple-output (MIMO) architecture has been introduced, wherein a very large number of distributed access points (APs) simultaneously and jointly serve a much smaller number of mobile stations (MSs). The paper extends the CF approach to the case in which both the APs and the MSs are equipped with multiple antennas, proposing a beamfoming scheme that, relying on the zero-forcing strategy, does not require channel estimation at the MSs. We contrast the originally proposed formulation of CF massive MIMO with a user-centric (UC) approach wherein each MS is served only by a limited number of APs. Exploiting the framework of successive lower-bound maximization, the paper also proposes and analyzes power allocation strategies aimed at either sum-rate maximization or minimum-rate maximization, both for the uplink and downlink. Results show that the UC approach, which requires smaller backhaul overhead and is more scalable that the CF deployment, also achieves generally better performance than the CF approach for the vast majority of the users, especially on the uplink.**

***Index Terms*— Cell-free massive MIMO, user-centric 5G cellular networks, uplink and downlink power allocation.**

I. Introduction

MASSIVE MIMO, introduced by Marzetta in his pioneering paper [1] is a promising 5G wireless access technology that can provide high throughput with simple signal processing [2]. Massive antenna at the base stations can be deployed in co-located or distributed setups. In co-located massive MIMO all the antennas are located in a compact area and this architecture has the advantage of low backhaul requirements. In distributed massive MIMO systems, instead, the antennas are spread out over a large area; this architecture has the advantage of efficiently exploiting mascroscopic diversity against the shadow fading, so these systems can potentially offer much higher probability of coverage than co-located massive MIMO [3], at the cost of increased backhaul requirements. Additionally, the distributed layout permits alleviating the cell-edge problem, since it considerably lowers the probability that a user happens to be situated far from every AP, permitting thus to achieve a better fairness and service uniformity across mobile users. In [4], the viability of using distributed antennas in multi-cell systems for massive MIMO on the uplink is investigated for a particular spatial correlation channel model. In [5] the authors focus on the downlink of a multicell distributed antenna system assuming that only the slowly-varying large-scale channel state is required at the transmitter and they explore the performance gain that can be achieved by coordinated transmissions for a virtual MIMO system. One of the drawbacks of such virtual MIMO systems is the heavy backhaul requirements, since, besides data symbols, also the channel estimates and the beamforming schemes are to be shared with the central processing unit (CPU).

Recently, a new architecture, named cell-free (CF) massive MIMO, has been introduced [6], [7], where a very large number of distributed single-antenna APs, whose ensemble forms a distributed massive MIMO array, serve many single-antenna MSs in the same time-frequency resource. All APs are connected to a CPU and cooperate via a backhaul network, serving all the MSs via time-division duplex (TDD) operation. In a CF massive MIMO system there are actually no cells or cell boundaries, and the system is such that the backhaul is used to transmit data-symbols on the downlink and sufficient statistics to the CPU on the uplink, to enable uplink data detection; otherwise stated, channel estimates at the APs are not forwarded to the CPU and the beamformers are computed locally. The CF concept thus can be interpreted as a scalable and lower-complexity implementation of distributed massive MIMO or of ultra-dense AP deployments with cloud-based radio access networks. Indeed, the distinguishing features of the CF user-centric architecture are the following: (a) the time division duplex protocol is used to exploit channel reciprocity

Manuscript received July 31, 2018; revised April 7, 2019, July 24, 2019, and September 19, 2019; accepted November 3, 2019. Date of publication November 14, 2019; date of current version February 11, 2020. This work was supported by the Italian Ministry of Education and Research through the program Dipartimenti di Eccellenza 2018–2022. This article was presented in part at the 21th International ITG Workshop on Smart Antennas, Berlin, Germany, March 2017, and in part at the 28th Annual IEEE International Symposium on Personal, Indoor and Mobile Radio Communications, Montreal, Canada, October 2017. The associate editor coordinating the review of this article and approving it for publication was R. Hu. *(Corresponding author: Stefano Buzzi.)*

S. Buzzi is with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, I-03043 Cassino, Italy, with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), I-43124 Parma, Italy, and also with GBB Wireless Research, I-80143, Napoli, Italy (e-mail: buzzi@unicas.it).

C. D'Andrea is with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, I-03043 Cassino, Italy, and also with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT), I-43124 Parma, Italy (e-mail: carmen.dandrea@unicas.it).

A. Zappone and C. D'Elia are with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio, I-03043 Cassino, Italy (e-mail: alessio.zappone@unicas.it; delia@unicas.it).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2019.2952117

on the uplink and downlink; (b) uplink channel estimates are computed locally at each AP and exploited locally, which means that they are not sent on the backhaul link; (c) beamformers to be used at the APs are computed locally and not at the CPU; (d) the backhaul is used to send data symbols on the downlink and sufficient statistics on the uplink to perform centralized uplink data decisions; however, it is not used to share the channel estimates and/or the beamformers.

CF massive MIMO systems have received increasing attention in the recent past. The authors of [7] show that the CF approach provides better performance than a small-cell system in terms of 95%-likely per-user throughput, thus confirming that the scheme is effective in alleviating the cell-edge user problem and in providing a more uniform service across users. The paper [8] has shown that some performance improvement can be obtained in low density networks by using downlink pilots, while [9], instead, analyzes the performance improvements granted by the use of a zero-forcing precoder in the downlink: although the gains are from five to ten-fold, the zero-forcing precoder requires centralized computations at the CPU and increased backhaul overhead. Zero-forcing precoding is again considered in [10], wherein it is coupled with a power control algorithm aimed at maximizing the energy efficiency of CF massive MIMO considering the backhaul power consumption and the imperfect channel state information (CSI). In [11], the uplink performance of CF systems is investigated using minimum mean square error (MMSE) processing for the case in which the energy efficiency of CF massive MIMO is to be maximized considering the backhaul power consumption and the imperfect CSI. The energy-efficiency of CF massive MIMO systems is also considered in [12], which proposes a power allocation algorithm aiming at maximizing the total energy efficiency, subject to a per-user spectral efficiency constraint and a per-AP power constraint. The power allocation strategies here are simplified by the fact that single-antenna transceivers are considered both at the APs and at the MSs, which permit skipping $\log\det[\cdot]$ functions in the achievable rate formulas. In order to reduce the backhaul overhead, the paper [13] considers instead a coded CF massive MIMO system, and investigates the performance of a compute & forward mechanism for the uplink, wherein each AP attempts to use an integer linear combination of the codewords to represent the scaled received signal to be sent to the CPU. The finite backhaul capacity is also considered in [14], which studies the case in which a quantized version of the estimated channel and the quantized received signal are available at the CPU, and the case when only the quantized version of the combined signal with maximum ratio combining detector is available at the CPU.

It should be noted that all the cited papers consider the case in which both the APs and the MSs are equipped with a single-antenna, the only exception being references [12], [14], which consider multiple antennas at the APs. The extension of the CF massive MIMO architecture to the case in which also the MSs are equipped with multiple antennas is not trivial since no channel estimation is performed at the MSs, and so no channel-dependent beamforming scheme can be used there. One critical point of the originally formulated version of the CF massive MIMO systems [6], [7] is the fact that all the APs serve all the MSs in the system. This assumption may lead to some inefficiencies in the system as the size of the considered area grows: indeed, it appears clearly pointless to waste power and computational resources at an AP to decode MSs that are very far and that are presumably received with a very low Signal-to-Interference plus Noise Ratio (SINR). To overcome this limitation, a UC distributed massive MIMO system has been introduced, still for single-antenna APs and MSs, in [15]; in the UC approach, each MS is served not by all the APs in the system, but just by the ones that are in the neighborhood. Similar APs selection strategies, based on either received power or largest large-scale fading, and exploiting a partial knowledge of the channel statistic, are proposed in reference [12]. The UC approach, while being much simpler than the CF one and less hungry of backhaul bandwidth, was shown in [15] to provide a larger achievable rate-per-user to the majority of the MSs in the system.

Paper contribution

Following on such a track, and building upon the conference papers [16], [17], this paper provides a thorough comparison of the UC and CF approaches, considering the case in which the MSs and the APs are equipped with multiple antennas. The paper focuses, in continuity with all the previously cited papers, on the case of sub-6 GHz frequencies, leaving for future investigations the case of higher carrier frequencies (see [18]–[20] for preliminary results on this). In the UC approach, it is assumed that each AP communicates only with a pre-assigned number of MSs, the ones that it receives with the strongest power. A zero-forcing beamforming scheme that does not require channel estimation at the MSs is proposed for use in the APs. This is in sharp contrast with the vast majority of the literature on massive MIMO, which assumes that the MSs are equipped with one antenna only. Channel estimation algorithms based on pilot matched (PM) and linear minimum mean square error (MMSE) criteria are developed. Lower bounds to the system achievable rate, taking into account the effect of channel estimation errors, are derived, and two power allocation strategies for the uplink and the downlink, are proposed. The former power allocation strategy maximizes the proposed lower bound for the system sum-rate, while the latter, targeting fairness across users, maximizes the minimum spectral efficiency lower bound across users. Both the optimization problems have non-concave objective functions, which makes their solution challenging; to solve them, the *successive lower-bound maximization* approach, merging the tools of *alternating optimization* and of *sequential convex programming*, is pursued. The numerical results will show that the proposed UC strategy is capable of outperforming the CF approach for the vast majority of the users in the network.

The remainder of this paper is organized as follows. Next Section contains the description of the considered system model. Section III is devoted to the illustration of the comunication protocol, composed by uplink training, downlink data transmission and uplink data transmission, for both CF and UC approaches. In Section IV the achievable rate expressions,

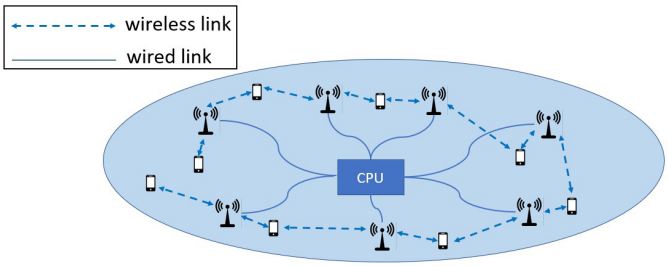

Fig. 1. UC approach to CF massive MIMO system.

and its lower bound, are derived for both downlink and uplink. In Section V we report the two power control strategies proposed for the downlink and Section VI contains the two power control strategies proposed for the uplink. Section VII contains the numerical results and finally, concluding remarks are given in Section VIII.

Notation: In this paper we use the following notation: $\mathbf{A}$ is a matrix; $\mathbf{a}$ is a vector; a is a scalar. The operators $(\cdot)^T$, $(\cdot)^{-1}$, and $(\cdot)^H$ stand for transpose, inverse and, conjugate transpose, respectively. The determinant of the matrix $\mathbf{A}$ is denoted as $|\mathbf{A}|$ and $\mathbf{I}_P$ is the $P \times P$ identity matrix. The vectorization operator is denoted by $\mathrm{vec}(\cdot)$ and the Kronecker product is denoted by $\otimes$. The statistical expectation operator is denoted as $\mathbb{E}[\cdot]$; $\mathcal{CN}\left(\mu, \sigma^2\right)$ denotes a complex circularly symmetric Gaussian random variable with mean μ and variance σ^2.

II. System Model

We consider an area with K MSs and M APs (see Fig. 1). MSs and APs are randomly located. The M APs are connected by means of a backhaul network to a CPU wherein data-decoding is performed. In keeping with the approach of [6], [7], all communications take place on the same frequency band; uplink and downlink are separated through time-division-duplex (TDD); the coherence interval is thus divided into three phases: (a) uplink channel estimation, (b) downlink data transmission, and (c) uplink data transmission. In phase (a) the MSs send pilot data in order to enable channel estimation at the APs. In phase (b) APs use channel estimates to perform channel-matched beamforming and send data symbols on the downlink; while in the CF architecture APs send data to all the MSs in the system, in the UC approach APs send data only to a subset of the MSs in the system. Finally, in phase (c) MSs send uplink data symbols to the APs; while in the CF architecture all the APs participate to the decoding of the data transmitted by all the MSs, in the UC approach APs just decode the data from the nearby MSs. The procedure for the selection of the MSs to serve will be specified in the following section. No pilots are transmitted on the downlink and no channel estimation is performed at the MSs: data decoding takes place on the downlink relying on the fact that in TDD the downlink channel is the reciprocal of the uplink channel.[1] In the following, we denote by N_{MS} and by N_{AP} the number of antennas at the MSs and at the APs, respectively.

[1]According to [7], the channel reciprocity is also ensured by perfect hardware chain calibration, whose feasibility has been recently shown in [21].

Before providing the detailed system mathematical model, it is worth noting that the UC approach can be seen as a special case of CF massive MIMO when specific power control rules are adopted. As an instance, a waterfilling-like downlink power control rule for a CF massive MIMO system may end up in a configuration where each AP serves only a limited number of MSs, thus resulting in an UC network deployment. On the other hand, CF can be also seen as a particular UC configuration, corresponding to the AP-MS association rule such that all the MS are associated to all the APs in the system.

Channel model

We denote by the $(N_{\mathrm{AP}} \times N_{\mathrm{MS}})$-dimensional matrix $\mathbf{G}_{k,m}$ the channel between the k-th MS and the m-th AP. We have

$$\mathbf{G}_{k,m} = \beta_{k,m}^{1/2} \mathbf{H}_{k,m}, \tag{1}$$

with $\beta_{k,m}$ a scalar coefficient modeling the channel shadowing effects and $\mathbf{H}_{k,m}$ an $(N_{\mathrm{AP}} \times N_{\mathrm{MS}})$-dimensional matrix whose entries are i.i.d $\mathcal{CN}(0,1)$ RVs. For the path loss and the shadow fading correlation models we use the ones reported in [7]. The large scale coefficient $\beta_{k,m}$ in (1) models the path loss and shadow fading, according to

$$\beta_{k,m} = 10^{\frac{\mathrm{PL}_{k,m}}{10}} 10^{\frac{\sigma_{\mathrm{sh}} z_{k,m}}{10}}, \tag{2}$$

where $\mathrm{PL}_{k,m}$ represents the path loss (expressed in dB) from the k-th MS to the m-th AP, and $10^{\frac{\sigma_{\mathrm{sh}} z_{k,m}}{10}}$ represents the shadow fading with standard deviation σ_{sh}, while $z_{k,m}$ will be specified later. For the path loss we use the following three slope path loss model [22]:

$$\mathrm{PL}_{k,m} = \begin{cases} -L - 35 \log_{10}\left(d_{k,m}\right), & \text{if } d_{k,m} > d_1 \\ -L - 10 \log_{10}\left(d_1^{1.5} d_{k,m}^2\right), & \text{if } d_0 < d_{k,m} \leq d_1 \\ -L - 10 \log_{10}\left(d_1^{1.5} d_0^2\right), & \text{if } d_{k,m} < d_0, \end{cases} \tag{3}$$

where $d_{k,m}$ denotes the distance between the m-th AP and the k-th user (expressed in km), while d_0 and d_1 are the breakpoint distances of the three slope path loss model. The quantity L is

$$\begin{aligned} L = {} & 46.3 + 33.9 \log_{10}(f) - 13.82 \log_{10}\left(h_{\mathrm{AP}}\right) \\ & - \left[1.11 \log_{10}(f) - 0.7\right] h_{\mathrm{MS}} + 1.56 \log_{10}(f) - 0.8, \end{aligned} \tag{4}$$

f is the carrier frequency in MHz, h_{AP} and h_{MS} denotes the AP and MS antenna heights, respectively. In real-world scenarios, transmitters and receivers that are in close vicinity of each other may be surrounded by common obstacles, and hence, the shadow fading RVs are correlated; for the shadow fading coefficient we thus use a model with two components [23]

$$z_{k,m} = \sqrt{\delta} a_m + \sqrt{1-\delta} b_k, \quad m = 1, \ldots, M, \ k = 1, \ldots, K, \tag{5}$$

where $a_m \sim \mathcal{N}(0,1)$ and $b_k \sim \mathcal{N}(0,1)$ are independent RVs, and δ, $0 \leq \delta \leq 1$ is a parameter. The covariance functions of a_m and b_k are given by:

$$E\left[a_m a_{m'}\right] = 2^{-\frac{d_{\mathrm{AP}(\mathrm{m},\mathrm{m}')}}{d_{\mathrm{decorr}}}} \qquad E\left[b_k b_{k'}\right] = 2^{-\frac{d_{\mathrm{MS}(\mathrm{k},\mathrm{k}')}}{d_{\mathrm{decorr}}}}, \tag{6}$$

where $d_{\mathrm{AP(m,m')}}$ is the geographical distance between the m-th and m'-th APs, $d_{\mathrm{MS(k,k')}}$ is the geographical distance between the k-th and the k'-th MSs. The parameter d_{decorr} is a decorrelation distance which depends on the environment, typically this value is in the range 20-200 m.

III. The Communication Protocol for the CF and UC Approaches

A. Uplink Training

We denote by τ_c the length (in samples) of the channel coherence time, and by τ_p the length (in samples) of the uplink training phase. Of course we must ensure that $\tau_p < \tau_c$. Denote by $\mathbf{\Phi}_k \in \mathcal{C}^{N_{\mathrm{MS}} \times \tau_p}$ the pilot sequence sent by the k-th MS, and assume that the rows of $\mathbf{\Phi}_k$ have unit norm. The signal received at the m-th AP in the n-th signaling time is represented by the following N_{AP}-dimensional vector:

$$\mathbf{y}_m(n) = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{G}_{k,m}\mathbf{\Phi}_k(:,n) + \mathbf{w}_m(n), \tag{7}$$

with p_k the k-th user transmit power on each antenna during the training phase. Collecting all the observable vectors $\mathbf{y}_m(n)$, for $n = 1, \ldots, \tau_p$ into the $(N_{\mathrm{AP}} \times \tau_p)$-dimensional matrix $\mathbf{Y}_m$, it is easy to show that:

$$\mathbf{Y}_m = \sum_{k=1}^{K} \sqrt{p_k}\mathbf{G}_{k,m}\mathbf{\Phi}_k + \mathbf{W}_m. \tag{8}$$

In the above equation the matrix $\mathbf{W}_m$ is $(N_{\mathrm{AP}} \times \tau_p)$-dimensional and contains the thermal noise contribution and out-of-cell interference at the m-th AP; its entries are assumed to be i.i.d. $\mathcal{CN}(0, \sigma_w^2)$ RVs. Based on the observable matrix $\mathbf{Y}_m$, the m-th AP performs estimation of the channel matrices $\{\mathbf{G}_{k,m}\}_{k=1}^{K}$.

1) PM Channel Estimation: For the PM channel estimation we assume knowledge of MSs transmit powers $\{p_k\}_{k=1}^{K}$. The estimate, $\widehat{\mathbf{G}}_{k,m}$ say, of the channel matrix $\mathbf{G}_{k,m}$ is obtained as

$$\widehat{\mathbf{G}}_{k,m} = \frac{1}{\sqrt{p_k}}\mathbf{Y}_m\mathbf{\Phi}_k^H = \mathbf{G}_{k,m}\mathbf{\Phi}_k\mathbf{\Phi}_k^H + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{\frac{p_j}{p_k}}\mathbf{G}_{j,m}\mathbf{\Phi}_j\mathbf{\Phi}_k^H + \frac{1}{\sqrt{p_k}}\mathbf{W}_m\mathbf{\Phi}_k^H. \tag{9}$$

Estimation (9) must be made in all the APs (i.e., for all the values of $m = 1, \ldots, M$) for all the values of $k = 1, \ldots, K$. If the rows of the matrices $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_K$ are pairwisely orthogonal (i.e. $\mathbf{\Phi}_k\mathbf{\Phi}_j = \mathbf{I}_{N_{\mathrm{MS}}}\delta_{i,k}$, for all i, k), then Eq. (9) simplifies to

$$\widehat{\mathbf{G}}_{k,m} = \frac{1}{\sqrt{p_k}}\mathbf{Y}_m\mathbf{\Phi}_k^H = \mathbf{G}_{k,m} + \frac{1}{\sqrt{p_k}}\mathbf{W}_m\mathbf{\Phi}_k^H, \tag{10}$$

and thermal noise is the only disturbance impairing the channel estimate. A necessary condition for this to happen is however $\tau_p \geq KN_{\mathrm{MS}}$, a relation that usually is not verified in practical scenarios due to the fact that τ_p must be a fraction of the channel coherence length. As a consequence, almost orthogonal pilot sequences are usually employed. In this paper, we assume that the pilot sequences assigned to each user are mutually orthogonal, so that $\mathbf{\Phi}_k\mathbf{\Phi}_k^H = \mathbf{I}_{N_{\mathrm{MS}}}$, while, instead, pilot sequences from different users are non-orthogonal. As a consequence, Eq. (9) is actually expressed as:

$$\widehat{\mathbf{G}}_{k,m} = \mathbf{G}_{k,m} + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sqrt{\frac{p_j}{p_k}}\mathbf{G}_{j,m}\mathbf{\Phi}_j\mathbf{\Phi}_k^H + \frac{1}{\sqrt{p_k}}\mathbf{W}_m\mathbf{\Phi}_k^H, \tag{11}$$

which clearly shows that the channel estimate is degraded not only by noise, but also by the pilots from the other users, an effect which is well-known to be named pilot contamination.

2) MMSE Channel Estimation: We assume now knowledge of the shadow fading coefficients $\beta_{k,m}, \forall\, m = 1, \ldots M$, $k = 1, \ldots K$ as in [6], [7]. We consider the $(N_{\mathrm{AP}}\tau_p)$-dimensional vector $\check{\mathbf{y}}_m = \mathrm{vec}\,(\mathbf{Y}_m)$, which can be shown to be expressed as

$$\check{\mathbf{y}}_m = \sum_{k=1}^{K} \sqrt{p_k}\left(\mathbf{\Phi}_k^T \otimes \mathbf{I}_{N_{\mathrm{AP}}}\right)\check{\mathbf{g}}_{k,m} + \check{\mathbf{w}}_m, \tag{12}$$

where $\check{\mathbf{g}}_{k,m} = \mathrm{vec}\,(\mathbf{G}_{k,m})$ is an $(N_{\mathrm{AP}}N_{\mathrm{MS}})$-dimensional vector and $\check{\mathbf{w}}_m = \mathrm{vec}\,(\mathbf{W}_m)$ is an $(N_{\mathrm{AP}}\tau_p)$-dimensional vector. Neglecting the correlation of the shadow fading, we assume $\mathbb{E}\left[\check{\mathbf{g}}_{k,m}\check{\mathbf{g}}_{k,m}^H\right] = \beta_{k,m}\mathbf{I}_{N_{\mathrm{AP}}N_{\mathrm{MS}}}$. Using the properties of the Kronecker product, the linear MMSE estimate of $\check{\mathbf{g}}_{k,m}$ can be shown to be obtained as [24]:

$$\widehat{\check{\mathbf{g}}}_{k,m} = \mathbf{R}_{k,m}\mathbf{R}_m^{-1}\check{\mathbf{y}}_m, \tag{13}$$

where $\mathbf{R}_{k,m} = \sqrt{p_k}\beta_{k,m}\left(\mathbf{\Phi}_k^* \otimes \mathbf{I}_{N_{\mathrm{AP}}}\right)$ and

$$\mathbf{R}_m = \sum_{j=1}^{K} p_j\beta_{j,m}\left(\mathbf{\Phi}_j^T\mathbf{\Phi}_j^* \otimes \mathbf{I}_{N_{\mathrm{AP}}}\right) + \sigma_w^2\mathbf{I}_{N_{\mathrm{AP}}\tau_p}. \tag{14}$$

B. Downlink Data Transmission

After the uplink channel estimation, the APs treat the channel estimates as the true channels, and a proper beamforming scheme is adopted, so as to ensure that the MSs will be able to receive data with no information on the channel state. Denoting by P_k the multiplexing order (i.e., the number of simultaneous data-streams) for user k, and by $\mathbf{x}_k^{\mathrm{DL}}(n)$ the P_k-dimensional unit-norm vector containing the k-th user data symbols to be sent in the n-th sample time, we let $\mathbf{L}_k = \mathbf{I}_{P_k} \otimes \mathbf{1}_{N_{\mathrm{MS}}/P_k}$ as the channel independent beamformer to be used at each MS. Basically, this corresponds to partitioning the MS antennas in as many disjoint subsets as the multiplexing order, and to use all the antennas in the same subset to transmit and receive one data stream. We stress that this channel independent beamforming at the MSs, based on its simple definition, remains the same in the two CF and UC approaches. The downlink precoder at the m-th AP for the k-th MS is expressed as

$$\mathbf{Q}_{k,m} = \widehat{\mathbf{G}}_{k,m}\left(\widehat{\mathbf{G}}_{k,m}^H\widehat{\mathbf{G}}_{k,m}\right)^{-1}\mathbf{L}_k. \tag{15}$$

It is easy to realize that using the above precoder at the AP, coupled with the use of the channel independent beamformer $\mathbf{L}_k$ at the MS, permits inverting the channel effect and perfectly recovering the transmitted symbols since we have that, in the ideal case of perfect channel knowledge, $\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} = \mathbf{I}_{P_k}$. Note also that computing the beamformers in (15) requires a computational effort proportional to $N_{\mathrm{AP}} N_{\mathrm{MS}}^3$, which is easily manageable since in the considered distributed environment, and for the case of sub-6 GHz frequencies, both N_{AP} and N_{MS} are small numbers.

1) CF Massive MIMO Architecture: In the CF architecture all the APs communicate with all the MSs in the systems, so the signal transmitted by the m-th AP in the n-th interval is the following N_{AP}-dimensional vector

$$\mathbf{s}_m^{\mathrm{cf}}(n) = \sum_{k=1}^{K} \sqrt{\eta_{k,m}^{\mathrm{DL,cf}}} \mathbf{Q}_{k,m} \mathbf{x}_k^{\mathrm{DL}}(n), \tag{16}$$

with $\eta_{k,m}^{\mathrm{DL,cf}}$ a scalar coefficient ruling the power transmitted by the m-th AP for the k-th MS. The generic k-th MS receives signal contributions from all the APs; the observable vector is expressed as

$$\begin{aligned} \mathbf{r}_k^{\mathrm{cf}}(n) &= \sum_{m=1}^{M} \mathbf{G}_{k,m}^H \mathbf{s}_m^{\mathrm{cf}}(n) + \mathbf{z}_k(n) \\ &= \sum_{m=1}^{M} \sqrt{\eta_{k,m}^{\mathrm{DL,cf}}} \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \mathbf{x}_k^{\mathrm{DL}}(n) \\ &\quad + \sum_{m=1}^{M} \sum_{\substack{j=1 \\ j \neq k}}^{K} \sqrt{\eta_{j,m}^{\mathrm{DL,cf}}} \mathbf{G}_{k,m}^H \mathbf{Q}_{j,m} \mathbf{x}_j^{\mathrm{DL}}(n) + \mathbf{z}_k(n). \end{aligned} \tag{17}$$

In (17), the N_{MS}-dimensional vector $\mathbf{z}_k(n)$, modelled as i.i.d. $\mathcal{CN}(0, \sigma_z^2)$ RVs, represents the thermal noise and out-of-cluster interference at the k-th MS. Based on the observation of the vector $\mathbf{r}_k^{\mathrm{cf}}(n)$, a soft estimate of the data symbols $\mathbf{x}_k^{\mathrm{DL}}(n)$ is obtained at the k-th MS as

$$\widehat{\mathbf{x}}_k^{\mathrm{DL,cf}}(n) = \mathbf{L}_k^H \mathbf{r}_k^{\mathrm{cf}}(n). \tag{18}$$

2) UC Massive MIMO Architecture: In the UC approach, we assume that the APs communicate only with the closest MSs. In order to define a measure for the closeness of the MSs, several procedures can be conceived. One possible strategy is that each AP computes the average Frobenius norm of the estimated channels for all the MSs, i.e.:

$$\bar{\mathbf{G}}_m = \frac{1}{K} \sum_{k=1}^{K} \| \widehat{\mathbf{G}}_{k,m} \|_F, \tag{19}$$

and will serve only the APs whose channel has a Frobenius norm larger than the computed average value. Another possible approach is that each AP sorts these estimates in descending Frobenius norm order and serves only the N MSs with the strongest channel, with N a proper design parameter. In this paper we will present numerical results using this latter strategy. We denote by $\mathcal{K}(m)$ the set of MSs served by the m-th AP. Given the sets $\mathcal{K}(m)$, for all $m = 1, \ldots, M$, we can define the set $\mathcal{M}(k)$ of the APs that communicate with the k-th user:

$$\mathcal{M}(k) = \{ m : k \in \mathcal{K}(m) \} \tag{20}$$

So, in this case, the signal transmitted by the m-th AP in the n-th interval is the following N_{AP}-dimensional vector

$$\mathbf{s}_m^{\mathrm{uc}}(n) = \sum_{k \in \mathcal{K}(m)} \sqrt{\eta_{k,m}^{\mathrm{DL,uc}}} \mathbf{Q}_{k,m} \mathbf{x}_k^{\mathrm{DL}}(n), \tag{21}$$

with $\eta_{k,m}^{\mathrm{DL,uc}}$, again, a scalar coefficient ruling the power transmitted by the m-th AP. The generic k-th MS receives signal contributions from all the APs; the observable vector is expressed as

$$\begin{aligned} \mathbf{r}_k^{\mathrm{uc}}(n) &= \sum_{m=1}^{M} \mathbf{G}_{k,m}^H \mathbf{s}_m^{\mathrm{uc}}(n) + \mathbf{z}_k(n) \\ &= \sum_{m \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL,uc}}} \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \mathbf{x}_k^{\mathrm{DL}}(n) \\ &\quad + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{m \in \mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL,uc}}} \mathbf{G}_{k,m}^H \mathbf{Q}_{j,m} \mathbf{x}_j^{\mathrm{DL}}(n) + \mathbf{z}_k(n). \end{aligned} \tag{22}$$

In (22), the N_{MS}-dimensional vector $\mathbf{z}_k(n)$ represents the thermal noise and out-of-cluster interference at the k-th MS, and is modeled as i.i.d. $\mathcal{CN}(0, \sigma_z^2)$ RVs. Based on the observation of the vector $\mathbf{r}_k^{\mathrm{uc}}(n)$, a soft estimate of the data symbols $\mathbf{x}_k^{\mathrm{DL}}(n)$ is obtained at the k-th MS as

$$\widehat{\mathbf{x}}_k^{\mathrm{DL,uc}}(n) = \mathbf{L}_k^H \mathbf{r}_k^{\mathrm{uc}}(n). \tag{23}$$

C. Uplink Data Transmission

The final phase of the communication protocol consists of the uplink data transmission. Since the MSs do not perform channel estimation, they just send their data symbols using the already defined trivial beamformer $\mathbf{L}_k$. We denote by $\mathbf{x}_k^{\mathrm{UL}}(n)$ the P_k-dimensional data vector to be transmitted by the k-th user in the n-th sample time. The signal received at the m-th AP in the n-th time sample is an N_{AP}-dimensional vector expressed as

$$\bar{\mathbf{y}}_m(n) = \sum_{k=1}^{K} \sqrt{\eta_k^{\mathrm{UL}}} \mathbf{G}_{k,m} \mathbf{L}_k \mathbf{x}_k^{\mathrm{UL}}(n) + \mathbf{w}_m(n), \tag{24}$$

with η_k^{UL} is the uplink transmit power of the k-th MS.

1) CF Massive MIMO Architecture: In the case of CF MIMO, all the APs participate to the decoding of the data sent by all the MSs. The m-th AP, thus, forms, for each $k = 1, \ldots, K$, the following statistics

$$\begin{aligned} \widetilde{\mathbf{y}}_{m,k}(n) &= \left(\mathbf{L}_k^H \widehat{\mathbf{G}}_{k,m}^H \widehat{\mathbf{G}}_{k,m} \mathbf{L}_k \right)^{-1} \mathbf{L}_k^H \widehat{\mathbf{G}}_{k,m}^H \bar{\mathbf{y}}_m(n) \\ &= \widetilde{\mathbf{G}}_{k,m} \bar{\mathbf{y}}_m(n), \end{aligned} \tag{25}$$

where we have defined $\widetilde{\mathbf{G}}_{k,m}$ as the following $P_k \times N_{\mathrm{AP}}$-dimensional matrix:

$$\widetilde{\mathbf{G}}_{k,m} = \left(\mathbf{L}_k^H \widehat{\mathbf{G}}_{k,m}^H \widehat{\mathbf{G}}_{k,m} \mathbf{L}_k \right)^{-1} \mathbf{L}_k^H \widehat{\mathbf{G}}_{k,m}^H. \tag{26}$$

The vectors $\widetilde{\mathbf{y}}_{m,k}(n)$, for all $k = 1, \ldots, K$, are then sent to the CPU via the backhaul link; the CPU, finally, forms the following soft estimates of the data vectors transmitted by the users:

$$\widehat{\mathbf{x}}_k^{\mathrm{UL,cf}}(n) = \sum_{m=1}^{M} \widetilde{\mathbf{y}}_{m,k}(n), \quad k = 1, \ldots, K. \tag{27}$$

Note that only the soft estimates $\widetilde{\mathbf{y}}_{m,k}(n)$ are to be transmitted from the APs to the CPU, while channel estimates transmission is not required.

2) UC Massive MIMO Architecture: In this case, the signal transmitted by the k-th MS is decoded only by the APs in the set $\mathcal{M}(k)$. Otherwise stated, the m-th AP computes the statistics $\widetilde{\mathbf{y}}_{m,k}(n)$ only for the MSs in $\mathcal{K}(m)$. Accordingly, the CPU is able to perform the following soft estimates for the data sent by the K MSs in the system:

$$\widehat{\mathbf{x}}_k^{\mathrm{UL,uc}}(n) = \sum_{m \in \mathcal{M}(k)} \widetilde{\mathbf{y}}_{m,k}(n), \quad k = 1, \ldots, K. \tag{28}$$

Notice that in this case the backhaul overhed is reduced with respect to the CF case since each AP has to send only the soft estimates of the data received by its associated MSs.

IV. Achievable Rate Expressions

We now focus on the derivation of achievable rate expressions. Since the CF approach can be obtained as a special case of the UC one by letting $N = K$, i.e. each AP serves all the K users in the system, so that $\mathcal{M}(k) = \{1, \ldots, M\}$, $\forall k = 1, \ldots, K$, in the following we focus on the UC case only and we omit the apex uc for simplicity of notation.

A. Perfect CSI

For benchmarking purposes, we start considering the ideal case in which perfect CSI is available both at the APs and at the MSs.

1) Downlink: From (23), we have that the achievable rate in downlink for the k-th user in the case of perfect CSI is written as

$$\mathcal{R}_{k,\mathrm{PCSI}}^{\mathrm{DL}} = W \log_2 \left| \mathbf{I}_{P_k} + \mathbf{R}_k^{-1} \mathbf{A}_{k,k} \mathbf{A}_{k,k}^H \right|, \tag{29}$$

where

$$\mathbf{A}_{k,j} = \sum_{m \in \mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL}}} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{G}_{j,m} \left(\mathbf{G}_{j,m}^H \mathbf{G}_{j,m} \right)^{-1} \mathbf{L}_j, \tag{30}$$

$$\mathbf{R}_k = \sigma_z^2 \mathbf{L}_k^H \mathbf{L}_k + \sum_{\substack{j=1 \\ j \neq k}}^{K} \mathbf{A}_{k,j} \mathbf{A}_{k,j}^H. \tag{31}$$

2) Uplink: From Eq. (28), we have that the achievable rate in uplink for the k-th user in the case of perfect CSI is

$$\mathcal{R}_{k,\mathrm{PCSI}}^{\mathrm{UL}} = W \log_2 \left| \mathbf{I}_{P_k} + \eta_k^{\mathrm{UL}} \widetilde{\mathbf{R}}_k^{-1} \mathbf{B}_{k,k} \mathbf{B}_{k,k}^H \right|, \tag{32}$$

where

$$\mathbf{B}_{k,j} = \sum_{m \in \mathcal{M}(k)} \left(\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{G}_{k,m} \mathbf{L}_k \right)^{-1} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{G}_{j,m} \mathbf{L}_j, \tag{33}$$

$$\widetilde{\mathbf{R}}_k = \sum_{\substack{j=1 \\ j \neq k}}^{K} \eta_j^{\mathrm{UL}} \mathbf{B}_{k,j} \mathbf{B}_{k,j}^H + \sigma_w^2 \sum_{m \in \mathcal{M}(k)} \left(\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{G}_{k,m} \mathbf{L}_k \right)^{-1}. \tag{34}$$

B. Imperfect CSI

In general, when the channel coefficients are not perfectly known, it is not clear what is "signal" and what is "interference" in Eqs. (22) and (28). In particular, for imperfect CSI, the intuitive notion of the rate in Eqs. (29) and (32) are in general not rigorously related to a corresponding notion of information theoretic achievable rate [25]. What can be instead done is to derive a lower bound (LB) to the achievable rate, as detailed in the following.

1) Downlink: In the following, we derive a LB for the achievable rate through the use-and-then-forget (UatF) bounding technique [26], [27]. We note that the received signal in (23) can be rewritten as

$$\begin{aligned} \widehat{\mathbf{x}}_k^{\mathrm{DL}}(n) = {} & \mathbb{E}\left[\sum_{m \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL}}} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \right] \mathbf{x}_k^{\mathrm{DL}}(n) \\ & + \left(\sum_{m \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL}}} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \right. \\ & \left. - \mathbb{E}\left[\sum_{m \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL}}} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \right] \right) \mathbf{x}_k^{\mathrm{DL}}(n) \\ & + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{m \in \mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL}}} \mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{j,m} \mathbf{x}_j^{\mathrm{DL}}(n) \\ & + \mathbf{L}_k^H \mathbf{z}_k(n), \end{aligned} \tag{35}$$

and an LB for the k-th MS downlink achievable rate can be written as

$$\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}) = W \log_2 \left| \mathbf{I}_{P_k} + \bar{\mathbf{R}}_k^{-1} \bar{\mathbf{A}}_k \bar{\mathbf{A}}_k^H \right|, \tag{36}$$

where $\bar{\mathbf{A}}_k = \sum_{m \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL}}} \bar{\mathbf{A}}_{k,m}$,

$$\bar{\mathbf{A}}_{k,m} = \mathbb{E}\left[\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \right], \tag{37}$$

$$\begin{aligned} \bar{\mathbf{R}}_k = {} & \sigma_z^2 \mathbf{L}_k^H \mathbf{L}_k + \sum_{m \in \mathcal{M}(k)} \sum_{\ell \in \mathcal{M}(k)} \sqrt{\eta_{k,m}^{\mathrm{DL}}} \sqrt{\eta_{k,\ell}^{\mathrm{DL}}} \mathbf{C}_{k,m,\ell} \\ & + \sum_{\substack{j=1 \\ j \neq k}}^{K} \sum_{m \in \mathcal{M}(j)} \sum_{\ell \in \mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL}}} \sqrt{\eta_{j,\ell}^{\mathrm{DL}}} \mathbf{E}_{k,j,m,\ell}, \end{aligned} \tag{38}$$

$$\begin{aligned} \mathbf{C}_{k,m,\ell} = {} & \mathbb{E}\left[\left(\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} - \mathbb{E}\left[\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{k,m} \right] \right) \right. \\ & \left. \times \left(\mathbf{L}_k^H \mathbf{G}_{k,\ell}^H \mathbf{Q}_{k,\ell} - \mathbb{E}\left[\mathbf{L}_k^H \mathbf{G}_{k,\ell}^H \mathbf{Q}_{k,\ell} \right] \right)^H \right], \end{aligned} \tag{39}$$

and

$$\mathbf{E}_{k,j,m,\ell} = \mathbb{E}\left[\mathbf{L}_k^H \mathbf{G}_{k,m}^H \mathbf{Q}_{j,m} \mathbf{Q}_{j,\ell}^H \mathbf{G}_{k,\ell} \mathbf{L}_k \right]. \tag{40}$$

From the above equations, it can be seen that the expression in Eq. (36) is deterministic and that it contains several expectations over the fast fading realizations. Notice that these expectations cannot be analytically computed, but, rather, can be numerically evaluated off-line through Monte Carlo simulations.

2) Uplink: Also for the uplink we use the UatF bounding technique. We rewrite Eq. (28) as:

$$\begin{aligned}\widehat{\mathbf{x}}_k^{\mathrm{UL}}(n) = {} & \mathbb{E}\left[\sum_{m\in\mathcal{M}(k)}\sqrt{\eta_k^{\mathrm{UL}}}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k\right]\mathbf{x}_k^{\mathrm{UL}}(n)\\ & + \left(\sum_{m\in\mathcal{M}(k)}\sqrt{\eta_k^{\mathrm{UL}}}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k\right.\\ & \left. - \mathbb{E}\left[\sum_{m\in\mathcal{M}(k)}\sqrt{\eta_k^{\mathrm{UL}}}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k\right]\right)\mathbf{x}_k^{\mathrm{UL}}(n)\\ & + \sum_{\substack{j=1\\ j\neq k}}^{K}\sum_{m\in\mathcal{M}(k)}\sqrt{\eta_j^{\mathrm{UL}}}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{j,m}\mathbf{L}_j\mathbf{x}_j^{\mathrm{UL}}(n)\\ & + \sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\mathbf{w}_m(n)\end{aligned} \tag{41}$$

The LB for the k-th MS uplink achievable rate can be written as

$$\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}\right) = W\log_2\left|\mathbf{I}_{P_k} + \eta_k^{\mathrm{UL}}\hat{\mathbf{R}}_k^{-1}\hat{\mathbf{A}}_k\hat{\mathbf{A}}_k^H\right|, \tag{42}$$

where $\hat{\mathbf{A}}_k = \mathbb{E}\left[\sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k\right]$,

$$\hat{\mathbf{R}}_k = \sigma_w^2\mathbf{G}_k + \eta_k^{\mathrm{UL}}\hat{\mathbf{C}}_k + \sum_{\substack{j=1\\ j\neq k}}^{K}\eta_j^{\mathrm{UL}}\hat{\mathbf{E}}_{k,j}, \tag{43}$$

$$\mathbf{G}_k = \mathbb{E}\left[\sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\widetilde{\mathbf{G}}_{k,m}^H\right], \tag{44}$$

$$\begin{aligned}\hat{\mathbf{C}}_k = {} & \mathbb{E}\left[\left(\sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k - \mathbb{E}\left[\sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{k,m}\mathbf{L}_k\right]\right)\right.\\ & \left.\times\left(\sum_{\ell\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,n}\mathbf{G}_{k,\ell}\mathbf{L}_k - \mathbb{E}\left[\sum_{\ell\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,\ell}\mathbf{G}_{k,\ell}\mathbf{L}_k\right]\right)^H\right],\end{aligned} \tag{45}$$

and

$$\hat{\mathbf{E}}_{k,j} = \mathbb{E}\left[\left(\sum_{m\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,m}\mathbf{G}_{j,m}\mathbf{L}_j\right)\left(\sum_{\ell\in\mathcal{M}(k)}\widetilde{\mathbf{G}}_{k,\ell}\mathbf{G}_{j,\ell}\mathbf{L}_j\right)^H\right]. \tag{46}$$

Also in this case the above expectations are hard to evaluate analytically, but can be computed through numerical simulations.

V. Downlink Power Control

We now propose power control algorithms optimizing the downlink achievable rate LBs, reported in (36). These algorithms are meant to be run in the system CPU, which is required to know the large-scale fading coefficients $\beta_{k,m}$, for all k and m. Let us denote by $\boldsymbol{\eta}$ the $KM\times 1$ vector collecting the downlink transmit powers of all APs for all MSs. We are thus concerned with the optimization of the downlink transmit powers for the maximization of the system sum-rate and minimum users' rate, subject to maximum power constraints. Mathematically, the sum-rate maximization problem is formulated as the optimization program:

$$\max_{\boldsymbol{\eta}} \sum_{k=1}^{K}\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}) \tag{47a}$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_m}\eta_{k,m}^{\mathrm{DL}} \leq P_{\max,m}, \quad \forall m = 1,\ldots,M \tag{47b}$$

$$\eta_{k,m}^{\mathrm{DL}} \geq 0, \quad \forall m = 1,\ldots,M,\ k = 1,\ldots,K, \tag{47c}$$

whereas the minimum rate maximization problem is

$$\max_{\boldsymbol{\eta}} \min_{1\leq k\leq K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}) \tag{48a}$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_m}\eta_{k,m}^{\mathrm{DL}} \leq P_{\max,m}, \quad \forall m = 1,\ldots,M \tag{48b}$$

$$\eta_{k,m}^{\mathrm{DL}} \geq 0, \quad \forall m = 1,\ldots,M,\ k = 1,\ldots,K. \tag{48c}$$

Both problems have non-concave objective functions, which makes their solution challenging. Moreover, even if the problems were concave, the large number of optimization variables, KM, would still pose a significant complexity challenge.[2] In order to face these issues, we will resort to the framework of successive lower-bound maximization, recently introduced in[3] [29], and briefly reviewed next.

A. Successive Lower-Bound Maximization

The main idea of the method is to merge the tools of alternating optimization [30, Section 2.7] and sequential convex programming [31]. To elaborate, consider the generic optimization problem

$$\max_{\boldsymbol{x}\in\mathcal{X}} f(\boldsymbol{x}), \tag{49}$$

with $f:\mathbb{R}^n\to\mathbb{R}$ a differentiable function, and $\mathcal{X}$ a compact set. As in the alternating optimization method, the successive lower-bound maximization partitions the variable space into M blocks, $\boldsymbol{x} = (\boldsymbol{x}_1,\ldots,\boldsymbol{x}_M)$, which are cyclically optimized one at a time, while keeping the other variable blocks fixed. This effectively decomposes (49) into M subproblems, with the generic subproblem stated as

$$\max_{\boldsymbol{x}_m} f(\boldsymbol{x}_m,\boldsymbol{x}_{-m}), \tag{50}$$

[2]Although polynomial, the best known upper-bound for the complexity of generic convex problems scales with the fourth power of the number of variables, while many classes of convex problems admit a cubic complexity [28].

[3]In [29] the method is labeled successive upper-bound minimization, since minimization problems are considered.

with $\boldsymbol{x}_{-m}$ collecting all variable blocks except the m-th. It is proved in [30, Proposition 2.7.1] that iteratively solving (50) monotonically improves the value of the objective of (49), and converges to a first-order optimal point if the solution of (50) is unique for any m, and if $\mathcal{X} = \mathcal{X}_1 \times \ldots \times \mathcal{X}_M$, with $\boldsymbol{x}_m \in \mathcal{X}_m$ for all m.

Clearly, alternating optimization proves useful when (50) can be solved with minor complexity. If this is not the case, the successive lower-bound maximization method proposes to tackle (50) by means of sequential convex programming. This does not guarantee to globally solve (50), but can lead to a computationally feasible algorithm. Moreover, it is guaranteed to preserve the properties of the alternating optimization method [29]. The idea of sequential optimization is to tackle a difficult maximization problem by solving a sequence of easier maximization problems. To elaborate, let us denote by $g_i(\boldsymbol{x}_m)$ the i-th constraint of (50), for $i = 1, \ldots, C$. Then, consider a sequence of approximate problems $\{\mathcal{P}_\ell\}_\ell$ with objectives $\{f_\ell\}_\ell$ and constraint functions $\{g_{i,\ell}\}_{i=1}^C$, such that the following three properties are fulfilled, for all ℓ:

- **(P1)** $f_\ell(\boldsymbol{x}_m) \leq f(\boldsymbol{x}_m)$, $g_{i,\ell}(\boldsymbol{x}_m) \leq g_i(\boldsymbol{x}_m)$, for all i and $\boldsymbol{x}_m$;
- **(P2)** $f_\ell(\boldsymbol{x}_m^{(\ell-1)}) = f(\boldsymbol{x}_m^{(\ell-1)})$, $g_{i,\ell}(\boldsymbol{x}_m^{(\ell-1)}) = g_i(\boldsymbol{x}_m^{(\ell-1)})$ with $\boldsymbol{x}_m^{(\ell-1)}$ the maximizer of $f_{\ell-1}$;
- **(P3)** $\nabla f_\ell(\boldsymbol{x}_m^{(\ell-1)}) = \nabla f(\boldsymbol{x}_m^{(\ell-1)})$, $\nabla g_{i,\ell}(\boldsymbol{x}_m^{(\ell-1)}) = \nabla g_i(\boldsymbol{x}_m^{(\ell-1)})$.

In [31] (see also [29], [32]) it is shown that, subject to constraint qualifications, the sequence $\{f(\boldsymbol{x}_m^{(\ell)})\}_\ell$ of the solutions of the ℓ-th Problem $\mathcal{P}_\ell$, is monotonically increasing and converges. Moreover, every convergent sequence $\{\boldsymbol{x}_m^{(\ell)}\}_\ell$ attains a first-order optimal point of the original Problem (50). Thus, the sequential approach enjoys strong optimality properties, fulfilling at the same time the monotonic improvement property for the objective function, and the Karush Kuhn Tucker (KKT) first-order optimality conditions for the original problem. Nevertheless, its applicability hinges on determining suitable lower bounds for the original objective to maximize, which fulfill all three properties **P1**, **P2**, **P3**, while at the same time leading to manageable optimization problems.

It is proved in [29] that successive lower-bound maximization has the same optimality properties as the true alternating optimization method, under similar assumptions, even though each subproblem might not be globally solved.[4]

B. Sum-Rate Maximization

Consider Problem (47) and define the variable blocks $\boldsymbol{\eta}_m$, $m = 1, \ldots, M$, collecting the transmit powers of AP m. Then, the sum-rate maximization with respect to the variable block $\boldsymbol{\eta}_m$ is cast as

$$\max_{\boldsymbol{\eta}_m} \sum_{k=1}^{K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{-m}) \tag{51a}$$

$$\text{s.t.} \sum_{k \in \mathcal{K}_m} \eta_{k,m}^{\mathrm{DL}} \leq P_{\max,m} \tag{51b}$$

$$\eta_{k,m}^{\mathrm{DL}} \geq 0, \quad \forall k \in \mathcal{K}_m. \tag{51c}$$

The complexity of (51) is significantly lower than that of (47), since only the M transmit powers of AP m are being optimized. Nevertheless, Problem (51) is still non-convex, which makes its solution difficult. Indeed, the k-th user's achievable rate can be expressed as (52) at the bottom of the next page, which can be seen to be non-concave, also with respect to only the variable block $\boldsymbol{\eta}_m$. Thus, following the successive lower-bound maximization approach, (51) will be tackled by sequential optimization. To this end, it is necessary to derive a lower-bound of the objective of (51), which fulfills Properties **P1**, **P2**, and **P3**, while at the same time leading to a simple optimization problem. To this end, the following lemma proves useful.

Lemma 1: The function $f : (x, y) \in \mathbb{R}^2 \to \sqrt{xy}$ is jointly concave in x and y, for $x, y > 0$.

Proof: The proof follows upon computing the Hessian of $\sqrt{xy}$ and showing that it is negative semi-definite. Details are omitted for the sake of brevity. □

Lemma 1, coupled with the facts that the function $\log_2|(\cdot)|$ is matrix-increasing, and that summation preserves concavity, implies that the rate function in (52), at the bottom of the next page, is the difference of two concave functions. This observation is instrumental for the derivation of the desired lower-bound. Indeed, recalling that any concave function is upper-bounded by its Taylor expansion around any given point $\boldsymbol{\eta}_{m,0}$, a concave lower-bound of $\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}$ is obtained as

$$\begin{aligned}\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}) &= g_1(\boldsymbol{\eta}_m) - g_2(\boldsymbol{\eta}_m) \\ &\geq g_1(\boldsymbol{\eta}_m) - g_2(\boldsymbol{\eta}_{m,0}) - \nabla_{\boldsymbol{\eta}_m}^T g_2|_{\boldsymbol{\eta}_{m,0}}(\boldsymbol{\eta} - \boldsymbol{\eta}_{m,0}) \\ &= \widetilde{\mathcal{R}}_k^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{m,0}).\end{aligned} \tag{53}$$

Moreover, it is easy to check that $\widetilde{\mathcal{R}}_k^{\mathrm{DL}}$ fulfills by construction also properties **P2** and **P3** with respect to $\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}$. Thus, Problem (51) can be tackled by the sequential optimization framework, by defining the ℓ-th problem of the sequence, $\mathcal{P}_\ell$, as the convex optimization program:

$$\max_{\boldsymbol{\eta}_m} \sum_{k=1}^{K} \widetilde{\mathcal{R}}_k^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{m,0}, \boldsymbol{\eta}_{-m}) \tag{54a}$$

$$\text{s.t.} \sum_{k \in \mathcal{K}_m} \eta_{k,m} \leq P_{\max,m} \tag{54b}$$

$$\eta_{k,m} \geq 0, \quad \forall k \in \mathcal{K}_m \tag{54c}$$

For any $\boldsymbol{\eta}_{m,0}$, Problem (54) can be solved by means of standard convex optimization theory, since the objective is concave, and the constraints are affine. The resulting power control procedure can be stated as in Algorithm 1. Moreover, based on the general theory reviewed in Section V-A, the following result holds.

Proposition 1: After each iteration in Line 6 *of Algorithm 1, the sum-rate value* $\sum_{k=1}^{K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}$ *is not decreased, and the resulting sequence* $\left\{\sum_{k=1}^{K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}\right\}$ *converges. Moreover, every*

[4]Of course, this holds provided the additional assumption of the sequential method are fulfilled in each iteration

Algorithm 1 Sum Rate Maximization

Set $i = 0$ and choose any feasible $\boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_M$;
repeat
  for $m = 1 \rightarrow M$ **do**
    repeat
      Choose any feasible $\boldsymbol{\eta}_{m,0}$;
      Let $\boldsymbol{\eta}_m^*$ be the solution of (54);
      $\boldsymbol{\eta}_{m,0} = \boldsymbol{\eta}_m^*$;
    until convergence
    $\boldsymbol{\eta}_m = \boldsymbol{\eta}_m^*$;
  end for
until convergence

limit point of the sequence $\{\boldsymbol{\eta}_m\}_m$ fulfills the KKT first-order optimality conditions of Problem (51).

Two remarks are now in order. First of all an extreme case of Algorithm 1 is that in which only one variable block is used, namely optimizing all of the transmit powers simultaneously. In this scenario, Algorithm 1 reduces to a pure instance of sequential optimization, and no alternating optimization is required. Nevertheless, as already mentioned, the complexity of this approach seems prohibitive for large M and K. Then, another extreme case is that in which the KM transmit powers $\eta_{k,m}$ are optimized one at a time, thus leading to considering KM variable blocks. The advantage of this approach is that each subproblem (54) would have only one optimization variable, and thus could be solved in semi-closed form. This brings drastic computational complexity savings and proves to be useful especially in the CF scenario, since in this case each variable block $\boldsymbol{\eta}_m$ always has dimension K.

1) Computational Complexity: The complexity of Algorithm 1 depends on the complexity of Problem (54), and on how many problems of the form of (54) must be solved before convergence.

As for (54), it is a convex problem and as such its complexity is polynomial in the number of variables, even though the specific degree of the polynomial is not known. The best available upper-bound for generic convex problems is provided in [28], and states that the complexity of any convex problem scales at most with the fourth power of the number of variables.

Instead, as for the number of iterations required for the "while" loops in Algorithm 1 to reach convergence, no closed-form result is currently available. Nevertheless, defining by I_O and I_S the number of iterations for the outer and inner "while" loops to converge, the overall complexity of Algorithm 1 can be upper-bounded by $\mathcal{O}\left(I_O I_S \sum_{m=1}^{M} |\mathcal{K}_m|^4\right)$.

C. Minimum Rate Maximization

Consider Problem (48). Following similar steps as in Section V-B, Problem (48) with respect to variable block $\boldsymbol{\eta}_m$ becomes

$$\max_{\boldsymbol{\eta}_m} \min_{1\le k\le K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{-m}) \tag{55a}$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_m} \eta_{k,m}^{\mathrm{DL}} \le P_{\max,m} \tag{55b}$$

$$\eta_{k,m}^{\mathrm{DL}} \ge 0, \quad \forall k \in \mathcal{K}_m \tag{55c}$$

Besides the difficulties already encountered in the sum-rate scenario, Problem (55) poses the additional challenge of having a non-differentiable objective due to the $\min(\cdot)$ operator. To circumvent this issue, (55) can be equivalently reformulated as the program:

$$\max_{\boldsymbol{\eta}_m, t} \; t \tag{56a}$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_m} \eta_{k,m}^{\mathrm{DL}} \le P_{\max,m} \tag{56b}$$

$$\eta_{k,m}^{\mathrm{DL}} \ge 0, \quad \forall k \in \mathcal{K}_m \tag{56c}$$

$$\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{-m}) \ge t, \quad \forall k = 1, \ldots, K. \tag{56d}$$

At this point, it is possible to tackle (56) by the sequential method. Leveraging again the bound in (53) leads to considering the approximate problem

$$\max_{\boldsymbol{\eta}_m, t} \; t \tag{57a}$$

$$\text{s.t.} \sum_{k\in\mathcal{K}_m} \eta_{k,m}^{\mathrm{DL}} \le P_{\max,m} \tag{57b}$$

$$\eta_{k,m}^{\mathrm{DL}} \ge 0, \quad \forall k \in \mathcal{K}_m \tag{57c}$$

$$\widetilde{\mathcal{R}}_k^{\mathrm{DL}}(\boldsymbol{\eta}_m, \boldsymbol{\eta}_{m,0}, \boldsymbol{\eta}_{-m}) \ge t, \quad \forall k = 1, \ldots, K. \tag{57d}$$

$$\begin{aligned}
&\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{DL}}(\boldsymbol{\eta}) \\
&= \underbrace{W \log_2 \left| \sigma_z^2 \mathbf{L}_k^H \mathbf{L}_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{m\in\mathcal{M}(j)} \sum_{\ell\in\mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL}} \eta_{j,\ell}^{\mathrm{DL}}} \mathbf{E}_{k,j,m,\ell} + \sum_{m\in\mathcal{M}(k)} \sum_{\ell\in\mathcal{M}(k)} \sqrt{\eta_{k,m}^{DL} \eta_{k,\ell}^{DL}} \left(\mathbf{C}_{k,m,\ell} + \bar{\mathbf{A}}_{k,m} \bar{\mathbf{A}}_{k,\ell}^H \right) \right|}_{g_1(\boldsymbol{\eta})} \\
&\quad - \underbrace{W \log_2 \left| \sigma_z^2 \mathbf{L}_k^H \mathbf{L}_k + \sum_{\substack{j=1 \\ j\neq k}}^{K} \sum_{m\in\mathcal{M}(j)} \sum_{\ell\in\mathcal{M}(j)} \sqrt{\eta_{j,m}^{\mathrm{DL}} \eta_{j,\ell}^{\mathrm{DL}}} \mathbf{E}_{k,j,m,\ell} + \sum_{m\in\mathcal{M}(k)} \sum_{\ell\in\mathcal{M}(k)} \sqrt{\eta_{k,m}^{DL} \eta_{k,\ell}^{DL}} \mathbf{C}_{k,m,\ell} \right|}_{g_2(\boldsymbol{\eta})}
\end{aligned} \tag{52}$$

Algorithm 2 Minimum Rate Maximization

Set $i = 0$ and choose any feasible $\boldsymbol{\eta}_2, \ldots, \boldsymbol{\eta}_M$;
repeat
  for $m = 1 \rightarrow M$ **do**
    repeat
      Choose any feasible $\boldsymbol{\eta}_{m,0}$;
      Let $\boldsymbol{\eta}_m^*$ be the solution of (57);
      $\boldsymbol{\eta}_{m,0} = \boldsymbol{\eta}_m^*$;
    until convergence
    $\boldsymbol{\eta}_m = \boldsymbol{\eta}_m^*$;
  end for
until convergence

For any $\boldsymbol{\eta}_{m,0}$, Problem (57) can be solved by means of standard convex optimization theory, since the objective is linear, and the constraints are all convex. The resulting power control procedure can be stated as in Algorithm 2, which enjoys similar properties as Algorithm 1.

1) Computational Complexity: Following similar arguments as done in Section V-B.1 for the sum-rate case, the complexity of Algorithm 2 is upper-bounded by $\mathcal{O}\left(I_O I_S \sum_{m=1}^{M} (|\mathcal{K}_m|+1)^4\right)$, where it has been accounted for the fact that the number of variables in the generic Problem (57) is $|\mathcal{K}_m|+1$ due to the presence of the auxiliary variable t.

VI. Uplink Power Control

Recalling (42) and denoting by $\widetilde{\boldsymbol{\eta}}$ the $K \times 1$ vector collecting the uplink transmit powers of all MSs, the uplink sum-rate maximization problem is stated as

$$\max_{\widetilde{\boldsymbol{\eta}}} \sum_{k=1}^{K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}\right) \tag{58a}$$

$$\text{s.t. } 0 \leq \eta_k^{\mathrm{UL}} \leq P_{\max,k} \quad \forall k = 1, \ldots, K, \tag{58b}$$

while the minimum rate maximization problem is stated as

$$\max_{\widetilde{\boldsymbol{\eta}}} \min_{1,\ldots,K} \mathcal{R}_{k,\mathrm{LB}}^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}\right) \tag{59a}$$

$$\text{s.t. } 0 \leq \eta_k^{\mathrm{UL}} \leq P_{\max,k} \quad \forall k = 1, \ldots, K. \tag{59b}$$

Proceeding similarly as for the downlink case, it is possible to develop power control algorithms for uplink sum-rate and minimum rate maximization, leveraging again the sequential optimization framework. Indeed, also in the uplink case, we observe that the k-th user's rate (42) can be written as the difference of two concave functions, as in Eq. (60) at the bottom of the next page. Now, it is clear that both $\widetilde{g}_1(\cdot)$ and $\widetilde{g}_2(\cdot)$ are concave functions of $\widetilde{\boldsymbol{\eta}}$ and thus (60) shows that the k-th user's rate can be once again written as the difference of two concave functions. As a consequence, for all $k = 1, \ldots, K$, a lower-bound of the k-th user's rate, which fulfills all three properties of the sequential optimization method, say $\widetilde{\mathcal{R}}_k^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}\right)$, is given by (53), in which $\widetilde{g}_1$ and $\widetilde{g}_2$ take the expressions in (60), at the bottom of the next page.

Remark 1: In the downlink case the number of optimization variables was KM, with $M > K$, and this made it convenient, for complexity reasons, to partition the variable space into multiple blocks of variables that were alternatively optimized. On the other hand, in the uplink case we only have K variables, and this makes it practically feasible to consider only one variable block, thus optimizing all variables at the same time. In the sequel, the focus will be on this case, but we stress that it, if desired, the optimization algorithms can be straightforwardly extended to the scenario in which multiple optimization blocks are defined and iteratively optimized.

Keeping Remark 1 in mind, both sum-rate maximization and minimum rate maximization can be performed by similar algorithms as Algorithms 1 and 2, respectively, in which the auxiliary problem to be solved within each iteration are stated as

$$\max_{\widetilde{\boldsymbol{\eta}}} \sum_{k=1}^{K} \widetilde{\mathcal{R}}_k^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\eta}}_0\right) \tag{61a}$$

$$\text{s.t. } 0 \leq \eta_k^{\mathrm{UL}} \leq P_{\max,k} \quad \forall k = 1, \ldots, K \tag{61b}$$

for sum-rate maximization, and as

$$\max_{\widetilde{\boldsymbol{\eta}}_m, t} \; t \tag{62a}$$

$$\text{s.t. } 0 \leq \eta_k^{\mathrm{UL}} \leq P_{\max,k} \quad \forall\, k = 1, \ldots, K \tag{62b}$$

$$\widetilde{\mathcal{R}}_k^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}, \widetilde{\boldsymbol{\eta}}_0\right) \geq t, \quad \forall\, k = 1, \ldots, K, \tag{62c}$$

for minimum rate maximization. Similar optimality properties as in the downlink case hold.

A. Computational Complexity

Following similar arguments as for the downlink scenario, the computational complexity of the proposed approach in the uplink scenario is upper-bounded by $\mathcal{O}\left(I_S K^4\right)$ in the sum-rate maximization case, and $\mathcal{O}\left(I_S (K+1)^4\right)$ in the minimum rate maximization case, where it has been accounted for the fact that, as stated in Remark 1, only one variable block is considered in the uplink scenario, thus removing the outer loop that is instead present in Algorithms 1 and 2 in the downlink scenario.

VII. Numerical Results

In our simulation setup, we consider a communication bandwidth of $W = 20$ MHz centered over the carrier frequency $f_0 = 1.9$ GHz. The antenna height at the AP is 15 m and at the MS is 1.65 m. The standard deviation of the shadow fading is $\sigma_{\mathrm{sh}} = 8$ dB, the parameters for the three slope path loss model in (3) are $d_1 = 50$ m and $d_0 = 10$ m, the parameter δ in (5) is 0.5 and the correlation distance in (6) is $d_{\mathrm{decorr}} = 100$ m. The additive thermal noise is assumed to have a power spectral density of -174 dBm/Hz, while the front-end receiver at the AP and at the MS is assumed to have a noise figure of 9 dB. In order to emulate an infinite area and to avoid boundary effects, the square area is wrapped around. The considered setup is taken from the originally formulated version of CF massive MIMO presented in reference [7]. The shown results come from an average over

TABLE I
PROBABILITY OF OBSERVING AN UNSERVED MS IN THE UC APPROACH FOR $M = 80$ AND $K = 15$

N	**1**	**2**	**3**	**4**	**5**	**6**	**7**
PCSI	0.12	0.028	0.008	0.0031	0.0014	$6\ 10^{-4}$	$2.8\ 10^{-4}$
MMSE CE	0.12	0.027	0.0086	0.0033	0.0015	$7\ 10^{-4}$	$3\ 10^{-4}$
PM CE	0.1	0.016	0.0028	$8\ 10^{-4}$	$2\ 10^{-4}$	10^{-4}	$1.3\ 10^{-5}$

N	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
PCSI	$1.2\ 10^{-4}$	$4.6\ 10^{-5}$	$2\ 10^{-5}$	$1.3\ 10^{-5}$	0	0	0	0
MMSE CE	$1.2\ 10^{-4}$	$5.3\ 10^{-5}$	$2\ 10^{-5}$	$6.6\ 10^{-6}$	0	0	0	0
PM CE	0	0	0	0	0	0	0	0

100 random scenario realizations with independent MSs and APs locations and channels. A square area of 1000×1000 (square meters) is considered, as in reference [7]; this area may be representative of a typical crowded environment where ultra-dense deployment of APs will be needed. We assume $N_{\mathrm{AP}} = 4$, $N_{\mathrm{MS}} = 2$ and the multiplexing order per user is $P_k = 2$, $\forall k = 1, \ldots, K$. During the uplink training phase, we use maximum-length-sequences (pseudo-noise) with length $\tau_p = 16$ and the uplink transmit power during the channel estimation for each antenna is $p_k = \tau_p \eta_k$, with $\eta_k = 100$ mW, $\forall k = 1, \ldots, K$. The considered power allocation rules will be compared for benchmarking with the uniform power allocation strategy. According to this strategy, in downlink each AP uniformly divides its maximum power among the users that it serves in the system. So, for the CF massive MIMO architecture we have

$$\eta_{k,m}^{\mathrm{DL,cf}} = \frac{P_{\max,m}}{K \mathrm{tr}\left(\mathbf{Q}_{k,m}\mathbf{Q}_{k,m}^H\right)}, \tag{63}$$

and for the UC massive MIMO architecture we have

$$\eta_{k,m}^{\mathrm{DL,uc}} = \begin{cases} \dfrac{P_{\max,m}}{\mathrm{card}\left[\mathcal{K}(m)\right] \mathrm{tr}\left(\mathbf{Q}_{k,m}\mathbf{Q}_{k,m}^H\right)} & \text{if } k \in \mathcal{K}(m) \\ 0 & \text{otherwise,} \end{cases} \tag{64}$$

where $\mathrm{card}\left[\mathcal{K}(m)\right]$ denotes the cardinality of the set $\mathcal{K}(m)$. For the uplink, instead, each MSs transmits with its maximum power, so in the CF and UC massive MIMO architecture we have

$$\eta_k^{\mathrm{UL,cf}} = \eta_k^{\mathrm{UL,uc}} = \frac{P_{\max,k}}{N_{\mathrm{MS}}}, \quad \forall k = 1, \ldots, K. \tag{65}$$

In the following results, we assume that the maximum power available at each AP is 200 mW, i.e. $P_{\max,m} = 200$ mW, $\forall\, m$ and the maximum power available at each MS is 100 mW, i.e. $P_{\max,k} = 100$ mW, $\forall k = 1, \ldots, K$.

We start by considering the outage probability in the UC approach. Indeed, one of the possible drawbacks in the considered AP-MS association rule is that it may happen that a MS does not end up associated with any AP. After that the association AP-MS has been made in the UC approach, possibly unserved MS might be associated to the closest AP to solve thus this problem. On the other hand, we show that the probability to have unserved MSs is very low, so having an unserved MS is a rare event. In Tables I and II the probability (estimated over 10^7 realizations) of having an unserved MS versus N is reported, for a high density and a low density scenario. In the high density scenario we have assumed $M = 80$ and $K = 15$, while in the low density scenario we have $M = 50$ and $K = 5$. We can see that increasing the value of N, in the cases of PCSI, MMSE channel estimation (CE) and PM CE, the outage probability decreases and reaches the value 0, i.e. all the users are served at least by one AP in the system. It can be also noted that in the case of PM CE the outage probability reaches the value 0 faster than in the case of PCSI; this can be due to the fact that the randomness introduced by the noise and the interference in the channel estimated reduces the probability of having an unserved MS.

TABLE II
PROBABILITY OF OBSERVING AN UNSERVED MS IN THE UC APPROACH FOR $M = 50$ AND $K = 5$

N	**1**	**2**	**3**	**4**	**5**
PCSI	0.048	0.005	$8\ 10^{-4}$	10^{-4}	0
MMSE CE	0.05	0.0058	0.001	10^{-4}	0
PM CE	0.021	$6\ 10^{-4}$	$4\ 10^{-5}$	0	0

We now focus on the evaluation of the achievable rate LB. Fig. 2 shows the cumulative distribution functions (CDFs) of the LB achievable rate per user in downlink for the CF and UC approaches for the case in which uniform power allocation (Uni) is used. Figs. 3 and Fig. 4 report the same plots as Fig. 2 with the only difference that the power control strategies maximizing the achievable sum-rate and and the minimum rate have been adopted, respectively. Again, the cases of PCSI, MMSE CE and PM CE. Regatrding the number of APs and of Ms, we consider an high density scenario, with $M = 80$,

$$\mathcal{R}_{k,\mathrm{LB}}^{\mathrm{UL}}\left(\widetilde{\boldsymbol{\eta}}\right) = \underbrace{W \log_2 \left| \sigma_w^2 \mathbf{G}_k + \sum_{\substack{j=1 \\ j \neq k}}^{K} \eta_j^{\mathrm{UL}} \hat{\mathbf{E}}_{k,j} + \eta_k^{\mathrm{UL}} \left(\hat{\mathbf{C}}_k + \hat{\mathbf{A}}_k \hat{\mathbf{A}}_k^H \right) \right|}_{\widetilde{g}_1(\widetilde{\boldsymbol{\eta}})} - \underbrace{W \log_2 \left| \sigma_w^2 \mathbf{G}_k + \sum_{\substack{j=1 \\ j \neq k}}^{K} \eta_j^{\mathrm{UL}} \hat{\mathbf{E}}_{k,j} + \eta_k^{\mathrm{UL}} \hat{\mathbf{C}}_k \right|}_{\widetilde{g}_2(\widetilde{\boldsymbol{\eta}})} \tag{60}$$

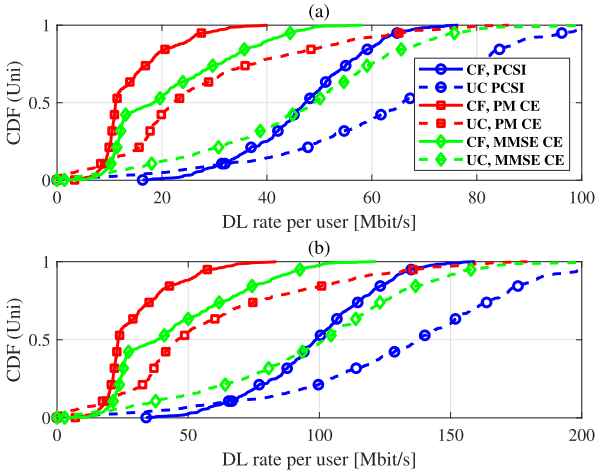

Fig. 2. CDF of rate per user LB in downlink with uniform power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

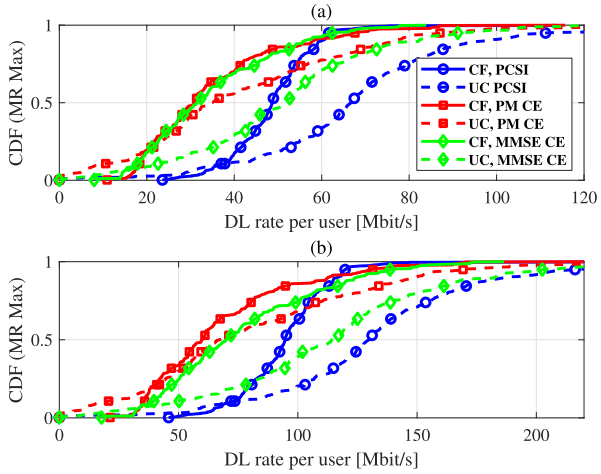

Fig. 4. CDF of rate per user LB in downlink with minimum-rate maximizing power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

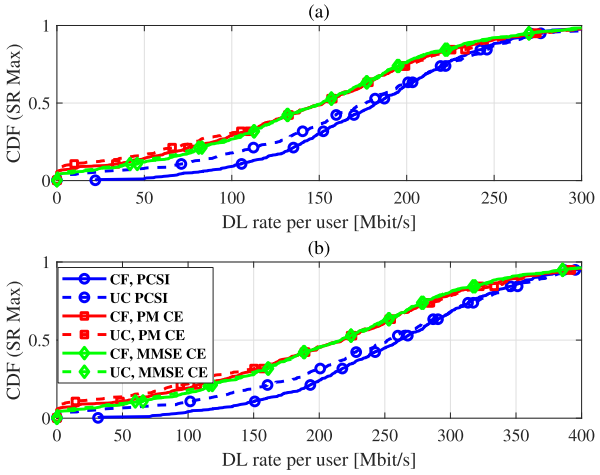

Fig. 3. CDF of rate per user LB in downlink with sum-rate maximizing power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

$K = 15$ and $N = 6$, and a low density scenario, with $M = 50$, $K = 5$ and $N = 2$. Inspecting these figures, the following comments can be offered.

- In the very low-rate region of the CDFs, the CF approach generally outperforms the UC approach; this part of the CDF reports the performance of the unlucky MSs that have very bad channels towards all the APs. These MSs take advantage of the CF deployment since they are served by a larger number of APs (compared to the UC case) and this explains the superiority of the CF approach for these MSs.
- The curves corresponding to UC and CF deployments usually cross, and, generally, UC approach outperforms the CF approach for the vast majority of users. This means that, excluding a small percentage of MSs with bas channel conditions, the UC approach is beneficial, probably due to the fact that each AP uses its power to transmit to nearby MSs and avoids wasting power in order to transmit to far MSs.
- This crossing does not take place in Fig. 3, wherein the performance corresponding to the maximum sum-rate power allocation policy is reported. This behavior can be explained by noticing that when power allocation is optimized, the CF approach provides a much greater flexibility than the UC case. Indeed, in the UC case, we optimize MN downlink transmit powers, forcing to zero the remaining $M(K - N)$ transmit powers; in the CF case, instead, we can optimize the system achievable sum-rate over MK transmit powers, so the solution space is much larger, and we have that from this point of view the UC approach can be seen as a special case of the CF one,[5] and this explain why for this case CF outperforms the UC approach.
- From Fig. 4, then, it is seen that, as expected, the CDF curves are steeper, since the power allocation maximizing the minimum rate introduces fairness among the MSs, so there should be no large disparities across the rates achieved by each MS.

Fig. 5 shows the average achievable system sum-rate LB in downlink versus N, assuming $M = 60$ and $K = 10$. From this figure, we can see that, again, if we focus on the sum-rate maximizing power allocation, the CF approach outperforms the UC one, whereas, if we consider the power allocation maximizing the minimum rate or the uniform power control allocations the UC achieves generally better performance, in terms of average achievable sum-rate, than the CF approaches. In the latter two cases, as far as system sum-rate is the performance measure, it is advisable to choose small values for N.

[5]This statement should not erroneously lead to the conclusion that the CF deployment is more general and provides better performance of the UC deployment. Indeed, while the statement only applies to downlink, it should also be considered that the CF deployment is not scalable and, also, the power allocation routine requires many iterations to converge given the larger number (with respect to UC approach) of variables to be optimized.

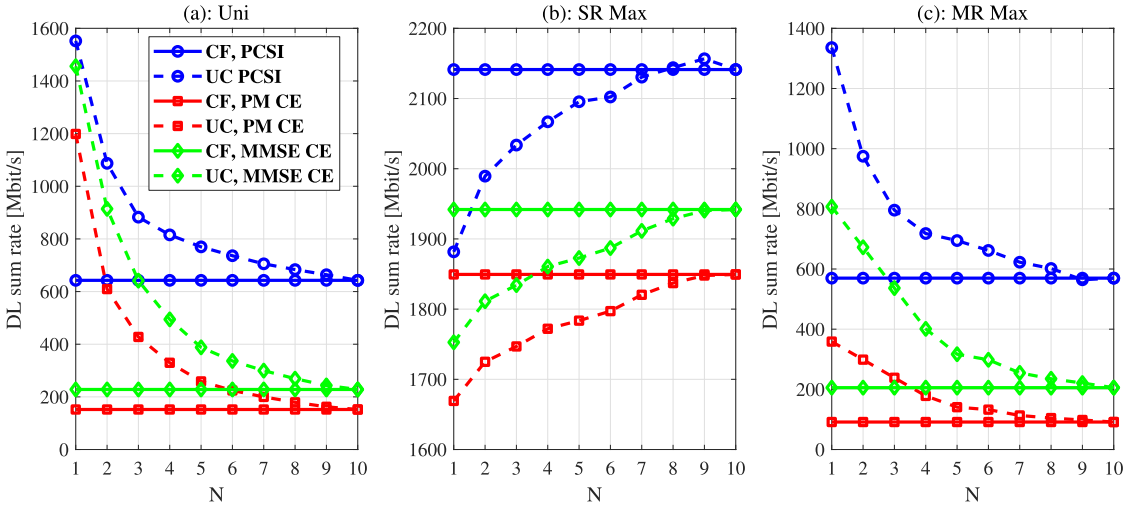

Fig. 5. System sum rate LB in downlink versus N. Parameters: $M = 60$, $K = 10$.

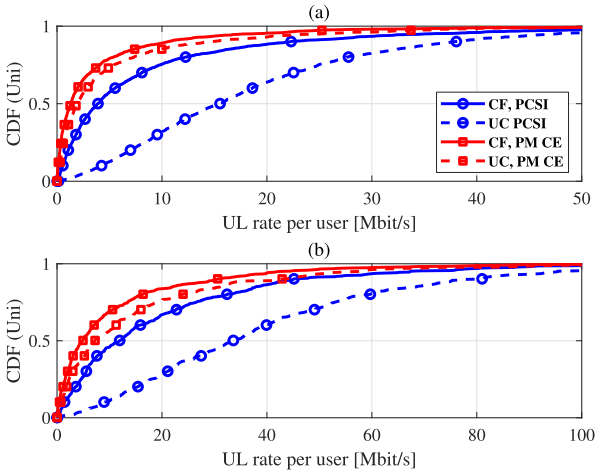

Fig. 6. CDF of rate per user LB in uplink with uniform power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

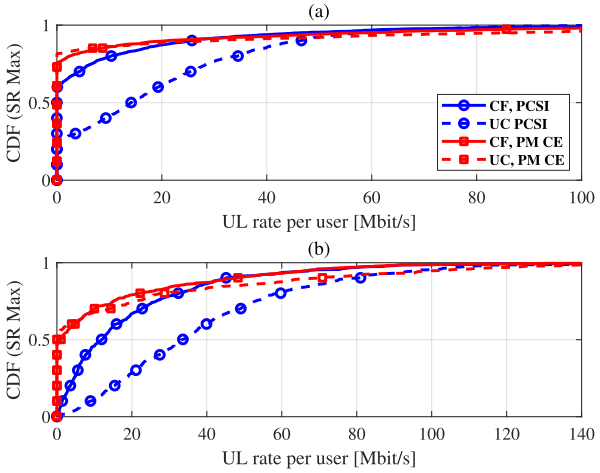

Fig. 7. CDF of rate per user LB in uplink with sum-rate maximizing power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

Let us now consider the uplink. In Fig. 6 we report the CDFs of the achievable rate LB per user for the CF and UC approaches for the case in which uniform power allocation

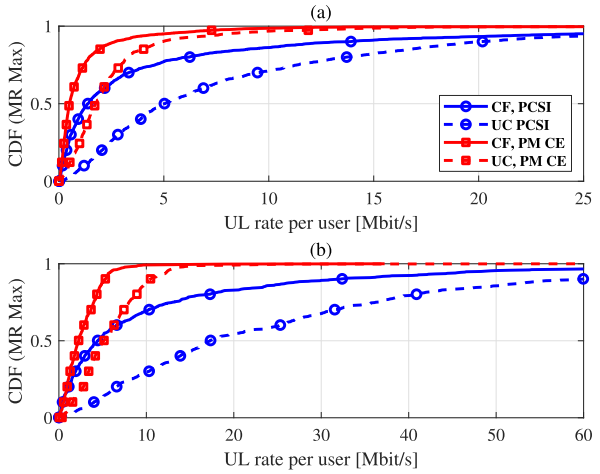

Fig. 8. CDF of rate per user LB in uplink with minimum-rate maximizing power allocation for a high density scenario in subfigure (a) and for a low density scenario in subfigure (b). Parameters: (a) $M = 80$, $K = 15$, $N = 6$; (b) $M = 50$, $K = 5$, $N = 2$.

is used. Figs. 7 and 8 show the same results for the case in which the power allocatio maximizes the system sum-rate and the minimum rate LBs, respectively. Again an high-density and a low-density scenario are considered. Fig. 9, finally, shows the system uplink sum-rate LB versus N, when the number of users is $K = 10$.

Inspecting these figures, we can see that, differently from the downlink, the UC approach outperforms the CF one in all the cases of uniform power allocation and power control strategies, and both in the cases of high and low density scenario. In particular, there are situations in which the UC approach guarantees many-fold improvements with respect to the CF strategy. This behavior can be explained by noting that, for the uplink, the UC can be no longer regarded as a special case of the CF configuration. In the uplink, the CF strategy requires that APs participate to the decoding of far MSs, and this adds a lot of additional noise to the decision statistic that ultimately endanger performance. In the UC approach, instead, each MS is decoded only by nearby APs, that can also

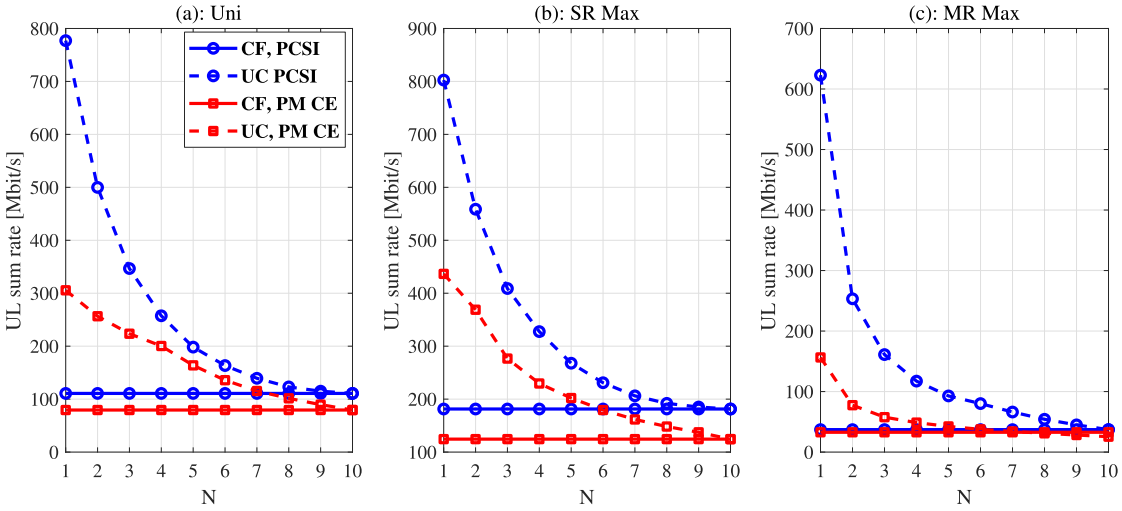

Fig. 9. System sum rate LB in uplink versus N. Parameters: $M = 60$, $K = 10$.

rely on good channel estimates, and this helps in considerably increasing the system performance.

VIII. Conclusion

The paper has focused on the recently introduced CF massive MIMO architecture. First of all, we have extended the CF approach to the case in which both the APs and the MSs are equipped with multiple antennas, and we have proposed the use of a channel-inverting beamforming scheme that does not require channel estimation at the MSs. Then, we have contrasted the CF architecture with the UC approach wherein each AP only decodes a pre-assigned number of MSs. We have proposed two power allocation strategies for the uplink and downlink, both for the CF and the UC cases. The first one is a sum-rate maximizing power allocation strategy, aimed at maximizing performance of the system in terms of overall data-rate, while the second one is a minimum-rate maximizing power allocation, aimed at maximizing performance of the system in terms of fairness. We compare the results of the power allocation strategies here proposed with the case of uniform power allocation. Results have shown that the UC approach generally outperforms the CF one, especially on the uplink. The UC approach thus exhibits in many relevant practical situations better performance than the CF approach, which motivates further investigation from the authors. Relevant research topics worth being investigated are, among others, the following: (a) the impact of user mobility on the performance of the CF and UC deployments; (b) the suitability of a UC architecture to support ultra-reliable low-latency communications; and, finally, (c) the coupling of CF massive MIMO architectures with 5G-and-beyond multiple access schemes such as the well-known non-orthogonal multiple access (NOMA). These topics form the object of current research.

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] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] S. Zhou, M. Zhao, X. Xu, J. Wang, and Y. Yao, "Distributed wireless communication system: A new architecture for future public wireless access," *IEEE Commun. Mag.*, vol. 41, no. 3, pp. 108–113, Mar. 2003.

[4] K. T. Truong and R. W. Heath, Jr., "The viability of distributed antennas for massive MIMO systems," in *Proc. 47th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2013, pp. 1318–1323.

[5] W. Feng, Y. Wang, N. Ge, J. Lu, and J. Zhang, "Virtual MIMO in multi-cell distributed antenna systems: Coordinated transmissions with large-scale CSIT," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 10, pp. 2067–2081, Oct. 2013.

[6] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO: Uniformly great service for everyone," in *Proc. 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 201–205.

[7] 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.

[8] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "How much do downlink pilots improve cell-free massive MIMO?" in *Proc. IEEE GLOBECOM*, Dec. 2016, pp. 1–7.

[9] E. Nayebi, A. Ashikhmin, T. L. Marzetta, H. Yang, and B. D. Rao, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[10] L. D. Nguyen, T. Q. Duong, H. Q. Ngo, and K. Tourki, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug. 2017.

[11] E. Nayebi, A. Ashikhmin, T. L. Marzetta, and B. D. Rao, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *Proc. 50th Asilomar Conf. Signals, Syst. Comput.*, Nov. 2016, pp. 203–207.

[12] 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.

[13] Q. Huang and A. Burr, "Compute-and-forward in cell-free massive MIMO: Great performance with low backhaul load," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 601–606.

[14] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–7.

[15] S. Buzzi and C. D'Andrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

[16] S. Buzzi and C. D'Andrea, "User-centric communications versus cell-free massive MIMO for 5G cellular networks," in *Proc. 21th Int. ITG Workshop Smart Antennas (WSA)*, Mar. 2017, pp. 1–6.

[17] S. Buzzi and A. Zappone, "Downlink power control in user-centric and cell-free massive MIMO wireless networks," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–6.

[18] M. Alonzo and S. Buzzi, "Cell-free and user-centric massive MIMO at millimeter wave frequencies," in *Proc. IEEE 28th Annu. Int. Symp. Pers., Indoor, Mobile Radio Commun. (PIMRC)*, Oct. 2017, pp. 1–5.
[19] M. Alonzo, S. Buzzi, and A. Zappone, "Energy-efficient downlink power control in mmWave cell-free and user-centric massive MIMO," in *Proc. IEEE 1st 5G World Forum (5GWF)*, Jul. 2018, pp. 493–496.
[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] F. Kaltenberger, H. Jiang, M. Guillaud, and R. Knopp, "Relative channel reciprocity calibration in MIMO/TDD systems," in *Proc. Future Netw. Mobile Summit*, Jun. 2010, pp. 1–10.
[22] A. Tang, J. Sun, and K. Gong, "Mobile propagation loss with a low base station antenna for NLOS street microcells in urban area," in *Proc. IEEE VTS 53rd Veh. Technol. Conf. VTC Spring*, vol. 1, May 2001, pp. 333–336.
[23] Z. Wang, E. K. Tameh, and A. R. Nix, "Joint shadowing process in urban peer-to-peer radio channels," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 52–64, Jan. 2008.
[24] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1998.
[25] G. Caire, "On the ergodic rate lower bounds with applications to massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 17, no. 5, pp. 3258–3268, May 2018.
[26] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.
[27] E. Björnson, J. Hoydis, and L. Sanguinetti, *Massive MIMO Networks: Spectral, Energy, and Hardware Efficiency*, vol. 11, nos. 3–4. Boston, MA, USA: Now, 2017.
[28] A. Ben-Tal and A. Nemirovski, *Lectures on Modern Convex Optimization: Analysis, Algorithms, and Engineering Applications*, vol. 2. Philadelphia, PA, USA: SIAM, 2001.
[29] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, Jun. 2013.
[30] D. P. Bertsekas, *Nonlinear Programming*. Belmont, MA, USA: Athena Scientific, 1999.
[31] B. R. Marks and G. P. Wright, "A general inner approximation algorithm for nonconvex mathematical programs," *Oper. Res.*, vol. 26, no. 4, pp. 681–683, Jul./Aug. 1978.
[32] A. Beck, A. Ben-Tal, and L. Tetruashvili, "A sequential parametric convex approximation method with applications to nonconvex truss topology design problems," *J. Global Optim.*, vol. 47, no. 1, pp. 29–51, May 2010.

Stefano Buzzi (M'98–SM'07) received the M.Sc. degree *(summa cum laude)* in electronic engineering and the Ph.D. degree in electrical and computer engineering from the University of Naples Federico II in 1994 and 1999, respectively. He had short-term research appointments at Princeton University, Princeton, NJ, USA, in 1999, 2000, 2001, and 2006. He is currently a Full Professor with the University of Cassino and Southern Lazio, Italy. He has coauthored about 160 technical peer-reviewed journals and conference papers, and among these, the highly-cited survey article "What will 5G be?" (IEEE JSAC, June 2014) on 5G wireless networks. His research interests are in the broad field of communications and signal processing, with an emphasis on wireless communications. He is also a member of the IEEE Future Networks Editorial Board. He serves regularly as a TPC member of several international conferences. He is also a former Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and the IEEE COMMUNICATIONS LETTERS. He has been the Lead Guest Editor of three IEEE JSAC special issues (June 2014, April 2016, and April 2019). He is also serving as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

Carmen D'Andrea (S'18) was born in Caserta, Italy, in July 1991. She received the B.S., M.S., and Ph.D. degrees (Hons.) in telecommunications engineering from the University of Cassino and Southern Lazio, Italy, in 2013, 2015, and 2019, respectively. In 2017, she was a Visiting Ph.D. Student with the Wireless Communications (WiCom) Research Group, Department of Information and Communication Technologies, Universitat Pompeu Fabra, Barcelona, Spain. She is currently a Post-Doctoral Researcher with the Department of Electrical and Information Engineering, University of Cassino and Southern Lazio. Her research interests are focused on wireless communication and signal processing, with a current emphasis on mmWave communications and massive MIMO systems, in both colocated and distributed setups.

Alessio Zappone received the M.Sc. and Ph.D. degrees from the University of Cassino and Southern Lazio, Cassino, Italy. In 2012, he has worked with the Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT) in the framework of the FP7 EU-funded project TREND. From 2012 to 2016, he was the PI with the Dresden University of Technology for the project CEMRIN on energy-efficient resource allocation in wireless networks, funded by the German Research Foundation (DFG). Since December 2019, he has been a tenured Professor with the University of Cassino and Southern Lazio. His research interests lie in the area of communication theory and signal processing, with a main focus on optimization techniques for resource allocation and energy efficiency maximization. He held several research appointments at international institutions. In 2017, he was a recipient of the H2020 Marie Curie IF BESMART Fellowship for experienced researchers, carried out at the LANEAS Group, CentraleSupelec, Gif-sur-Yvette, France. He was appointed as an Exemplary Reviewer for the IEEE TRANSACTIONS ON COMMUNICATIONS in 2016 and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS in 2017 and 2018. He also serves as a Senior Area Editor for the IEEE SIGNAL PROCESSING LETTERS. He has served as a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS ON COMMUNICATIONS (Special Issues on Energy-Efficient Techniques for 5G Wireless Communication Systems and on Wireless Networks Empowered by Reconfigurable Intelligent Surfaces).

Ciro D'Elia received the M.S. degree (Hons.) in telecommunication engineering and the Ph.D. degree in computer science and electronic engineering from the University of Naples Federico II, Naples, Italy, in 1998 and 2001, respectively. Since October 2001, he has been a Researcher with the Department of Electrical and Information Engineering (DIEI), University of Cassino and Southern Lazio, Cassino, Italy, where he teaches the M.S. courses on Telecommunication Networks, Telematics, Image Processing and Transmission, and the Ph.D. course on Object-Oriented Design for Signal Processing. Since 2002, he has been the Technical Director of the Informatics and Telecommunication Laboratory, DIEI. He has participated in several research projects funded by the Italian Ministry of Education University and Research (MIUR), the Italian Space Agency (ASI), the German Aerospace Center (DLR), and the European Space Agency (ESA). He held contracts from several national companies: Telecom Italia, Sogin, MBDA Italia, Urmet, and ACEA ATO2 and foreign companies: MetaSensing, The Netherlands, VicomTech, Spain. His research interests are in statistical signal and image processing and mining, information extraction from remotely sensed data, telecommunication networks, telematics, network security, the IoT, and cybersecurity.