

Towards Optimal Energy Efficiency in Cell-Free Massive MIMO Systems

A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, S. Chatzinotas, and J. M. Senior

Abstract—Motivated by the ever-growing demand for *green* wireless communications and the advantages of *cell-free* (CF) massive multiple-input multiple-output (mMIMO) systems, we focus on the design of their downlink (DL) for optimal *energy efficiency* (EE). To address this fundamental topic, we assume that each access point (AP) is deployed with multiple antennas and serves multiple users on the same time-frequency resource while the APs are Poisson point process (PPP) distributed, which approaches realistically their opportunistic spatial randomness. Relied on tools from stochastic geometry, we derive a lower bound on the DL average achievable spectral efficiency (SE). Next, we consider a realistic power consumption model for CF mMIMO systems. These steps enable the formulation of a tractable optimization problem concerning the DL EE, which results in the analytical determination of the optimal pilot reuse factor, the AP density, and the number of AP antennas and users that maximize the EE. Hence, we provide useful design guidelines for CF mMIMO systems relating to fundamental system variables towards optimal EE. Among the results, we observe that an optimal pilot reuse factor and AP density exist, while larger values result in an increase of the interference, and subsequently, lower EE. Overall, it is shown that the CF mMIMO technology is a promising candidate for next-generation networks achieving simultaneously high SE and EE.

Index Terms—Cell-free massive MIMO systems, energy efficiency, stochastic geometry, small cells networks, beyond 5G MIMO.

I. INTRODUCTION

The rapid development of wireless communication systems, by means of the fifth generation (5G) networks and beyond, aim at higher data rates with adequate quality of service (QoS) but with the reduction of energy consumption being of primary concern [2]. Obviously, achieving higher data rates with less power consumption might seem like contradictory goals [3], but that is not necessarily the case. A promising solution to provide higher data rates is achieved by means of the so-called network densification, which, unfortunately, stumbles

at the major bottleneck of increasing interference resulting in higher power consumption [4]. Hence, the fundamental arising question is how to increase the network data rate while achieving optimal energy efficiency (EE) at the same time. Although both academia and industry already have focused on the EE of cellular networks in the past years [5], existing architectures cannot face the increasing complexity of future networks towards many devices, many antennas, and many bands. For this reason, new innovative architectures are required to address the crucial demanding *green* specifications and considerations in next-generation networks.

In the direction of network densification, a key 5G technology (in terms of the number of antennas per area unit), known as massive multiple-input multiple-output (mMIMO) systems, has emerged by providing $10\times$ higher data rate with comparison to conventional cellular systems [6]–[9]. Although mMIMO systems can effectively deal with interference, the achievable EE is limited by the large propagation losses that are typical in cellular networks. An interesting alternative is to distribute a large number of antennas over the coverage area and operate these antennas in a network MIMO manner [10], [11]¹. A practical embodiment of network MIMO is the cell-free (CF) mMIMO concept described in [12].

CF mMIMO consist of a deployment of a large number of access points (APs) that are distributed over the coverage area to coherently serve a large number of users on the same time-frequency resource. According to [12], as the number of APs increases, we manage to take advantage of the favorable propagation and channel hardening properties, and finally, achieve very large spectral efficiency (SE) with simplified signal processing needing less overhead. However, herein, it is crucial to mention that the attractive properties of channel hardening and favorable propagation do not hold under all conditions. In particular, despite [12] that accounted for these properties for single-antenna APs, in [13], it was proved the opposite. Fortunately, it was shown that channel hardening and favorable propagation appear in the case of multiple-antennas APs (at least 5 – 10 antennas) or low path-loss. As a result, CF mMIMO can combine the benefits of coordination and low overhead. Moreover, CF mMIMO is a promising architecture because by increasing the number of APs the path-losses are improved and the macro-diversity is enhanced [12], which means that the transmit powers can be reduced. Unfortunately, these gains from CF mMIMO are achieved by deploying more hardware, which in turn, may increase the power consumption.

¹Despite that network MIMO has attracted a lot of interest in the last decade [10], [11], its implementation is not feasible for practical systems due to its substantial backhaul overhead.

A. Papazafeiropoulos is with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K., and with SnT at the University of Luxembourg, Luxembourg. H. Q. Ngo is with the School of Electronics, Electrical Engineering and Computer Science, Queen’s University Belfast, Belfast BT3 9DT, U.K. P. Kourtessis and John M. Senior are with the Communications and Intelligent Systems Research Group, University of Hertfordshire, Hatfield AL10 9AB, U. K. S. Chatzinotas is with the SnT at the University of Luxembourg, Luxembourg. E-mails: tapapazaf@gmail.com, hien.ngo@qub.ac.uk, p.kourtessis@herts.ac.uk, symeon.chatzinotas@uni.lu.

This work was supported in part by the University of Hertfordshire’s 5-year Vice Chancellor’s Research Fellowship and by the National Research Fund, Luxembourg, under the projects ECLÉCTIC and 5G-SKY. The work of H. Q. Ngo was supported by the U.K. Research and Innovation Future Leaders Fellowships under Grant MR/S017666/1. This paper was presented at the IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications, London, U.K., Sep. 2020 [1].

Notably, even though CF mMIMO systems come with plausible potentials, the study of this technology is limited as literature reveals [1], [12]–[23]. For example, the authors in [15] achieved better data rates by suggesting a user-centric approach of CF mMIMO systems, where the APs serve a group of users instead of all of them. Another interesting study concerns [1], where the locations of the APs are Poisson point process (PPP) distributed, and the coverage probability was derived².

Along the line concerning energy consumption, there are many works that have studied the EE of massive MIMO [25]–[27]. Specifically, in [25], the optimal uplink (UL) EE of cellular networks was obtained analytically and examined thoroughly by using tools from stochastic geometry with PPP distributed BSs, and in [27], the same methodology was applied for a multislope path-loss model. Only a few prior works have examined the EE in CF mMIMO systems which is of particular interest since they are more beneficial than network MIMO [14], [19], [28]. In particular, in the case of CF mMIMO systems, the EE was investigated in [14], [28] while, in [19], the EE was investigated under a user-centric approach at millimeter-wave frequencies, but these works did not obtain analytical expressions for the EE, and did not take spatial randomness into account.

A. Motivation

Most existing works on 5G networks focus on the SE while they neglect the importance of EE which is decreased when interference increases. Network MIMO, mitigating interference by means of coordination, is practically unattainable due to excessively high complexity in terms of hardware and information overhead. Luckily, CF mMIMO systems are an embodiment of massive MIMO and network MIMO systems and emerge as a promising feasible solution regarding coordination with low overhead exploiting the favorable propagation and channel hardening properties as the number of each AP antennas increases. Hence, the study of EE of CF mMIMO systems is of pivotal interest. Despite some existing works on the numerical optimization of the EE of CF mMIMO systems [14], [28], there is no previous work deriving the optimal system parameters in closed form. Most importantly, existing works, except [1], [17], focus on simplified network topologies such as grid-based models, and they do not account for the realistic spatial randomness of the APs³. Especially, as the number of APs increases according to the concept of CF mMIMO, they are deployed opportunistically which means high irregularity. Although previous works mentioned that the

²Note that PPP is the most popular and tractable point process for describing the spatial distribution of network nodes, e.g., see [24] and relevant works. The consideration of other spatial distributions such as the Matern Hard-core point process could be the topic of future research.

³In [17], the spatial randomness of the APs was considered. However, the distribution of the APs was again idealized and neglected their irregularity since it was assumed uniform, i.e., a binomial point process (BPP) was applied. Moreover, certain approximations were made that result in a not strict analysis with not reliable expressions. For example, it was made the assumption of the nearest AP and it was considered the mean contribution from the rest of the APs. Regarding our recent work in [1], it was relied on the deterministic equivalent (DE) analysis to obtain the DE signal-to-interference-plus-noise ratio (SINR) for a large number of APs. Also, it focused on the derivation of the coverage probability and achievable rate for a large APs number.

APs are randomly located, they consider a fixed number of APs while their randomness is not utilized in the analysis, but only in the simulations. These observations suggest that the analytical derivation of the optimal realistic EE of CF mMIMO systems, where the APs are distributed according to a PPP, is of paramount importance. Also, in order to extract trustworthy results, a realistic power consumption model is needed to take both the transmit power and other system parameters into account.

B. Contribution

The main contributions are summarized as follows.

- Contrary to existing works [14], [28], which did not account for the spatial randomness of the APs, and thus, are quite idealized, we apply tools from stochastic geometry and assume that the APs are PPP located. In addition, contrary to [1], our analysis relies on a finite number of APs, and the aim of this work is the study of the EE. Also, we differentiate from [17] that assumed a BPP for the APs which is again idealistic.
- We derive a lower bound of the downlink (DL) average achievable SE for a finite number of APs being PPP distributed and having multiple antennas. Furthermore, we present a realistic power consumption model, specialized in CF mMIMO systems.
- Contrary to the common definition of area SE (ASE) in cellular networks, we provide a novel definition, which is necessary for CF mMIMO systems, and in general, in architectures with coordinated multi-point joint transmission (CoMP-JT).
- We obtain the optimal EE of CF mMIMO systems with PPP distributed multiple-antenna APs by means of an analytical expression enabling to derive the optimal values for fundamental system parameters such as the network size in terms of numbers of AP antennas and serving users.
- We shed light on the impact of the main system parameters on the optimal EE. The results are of high practical interest since the analysis accounts for finite and realistic systems dimensions. Specifically, we obtain the optimal reuse factor, the optimal AP density, and the optimal number of AP antennas and users in closed-forms. For the sake of comparison, we also present results for a corresponding “cellular” mMIMO system and a small-cells (SCs) network.

C. Paper Outline

The remainder of this paper is organized as follows. Section II presents the system model of a CF mMIMO system with multiple antennas APs being PPP distributed. Sections III and IV provide the UL training and DL transmission phases, respectively. Section V provides the analysis regarding the EE while Section VI presents the optimization of the EE and obtains the optimal system parameters in closed form. The numerical results are placed in Section VII, and Section VIII concludes the paper.

D. Notation

Vectors and matrices are denoted by boldface lower and upper case symbols, respectively. The symbols $(\cdot)^\top$, $(\cdot)^H$, and $\text{tr}(\cdot)$ express the transpose, Hermitian transpose, and trace operators, respectively. The expectation operator is denoted by $\mathbb{E}[\cdot]$. Also, $\mathbf{b} \sim \mathcal{CN}(\mathbf{0}, \Sigma)$ represents a circularly symmetric complex Gaussian vector with zero mean and covariance matrix Σ . Finally, the superscript $*$ is used to represent optimal values.

II. SYSTEM MODEL

We consider a CF mMIMO system with multiple antennas at the APs and we model the practical spatial randomness of APs by means of stochastic geometry. Specifically, we assume that the APs, each having $N \geq 1$ antennas, are distributed in the two dimensional Euclidean plane with their locations following a homogeneous PPP Φ_{AP} with intensity λ_{AP} [AP/km²]. In a specific realization of the PPP Φ_{AP} , the number of APs in any region \mathcal{A} of size S in km², denoted by M , is a Poisson random variable with mean value

$$\mathbb{E}[M] = \lambda_{\text{AP}} S. \quad (1)$$

Following the network MIMO principle, all the APs serve simultaneously all the single-antenna users on the same time-frequency resource⁴. Interestingly, the total number of antennas in \mathcal{A} in a realization of the spatial process, denoted by $\mathcal{W} = MN$ is a Poisson random variable with mean $\mathbb{E}[\mathcal{W}] = N\lambda_{\text{AP}}S$. We let K denote the number of users in any given network realization. Their number is fixed and the users are selected at random from a large set based on round-robin scheduling⁵. Notably, the number of users is an optimization variable while their locations are uniformly distributed [25]. To consider a CF mMIMO scenario, the densities are chosen in order to fulfill the condition $\mathcal{W} \gg K$ in most realizations [13].

All APs are connected via a perfect fronthaul network to a central processing unit⁶. Taking advantage of Slivnyak's theorem, we focus on a typical user, selected at random among the users and indexed by k , in order to analyze the network performance [30]. For the ease of exposition, we assume that the typical user is located at the origin.

A. Channel Model

In a realization of the PPP Φ_{AP} , i.e., given M , let the $N \times 1$ channel vector \mathbf{h}_{mk} between the m th AP and the typical user be given by

$$\mathbf{h}_{mk} = l_{mk}^{1/2} \mathbf{g}_{mk}, \quad m = 1, \dots, M \text{ and } k = 1, \dots, K \quad (2)$$

⁴Given that our focus is the study of CF mMIMO systems under practical assumptions, the optimization of their EE by accounting for user-centric and scalable requirements as in [15] and [18] is a topic of future research.

⁵This choice is equivalent to a random user selection in each time-frequency resource block and it is a common assumption in the literature for analytical tractability. The study of the impact of optimal scheduling is an interesting topic for future research.

⁶Although the fronthaul links are not perfect in practice, but degraded due to several reasons such as the quantization noise [17], [29], this work assumes perfect fronthaul connections to focus on the impact of a realistic spatial randomness of the APs. The consideration of the fronthaul links limitations is of practical interest and is left for future work.

where $l_{mk} = \min(1, r_{mk}^{-\alpha})$ and \mathbf{g}_{mk} represent independent path-loss and small-scale fading between the m th N antenna AP and the typical user. In particular, the path-loss is described by means of a non-singular bounded model with $\alpha > 0$ being the path-loss exponent and r_{mk} being the distance between the m th AP and the k th user [31]. Note that this bounded path-loss model is practical also at short distances while an bounded path-loss model is not suitable for the study of CF mMIMO systems with stochastic geometry because it might result in unrealistically high power gain if the APs arbitrarily close to the user [13]. Given that this work accounts for the spatial randomness of the APs, the following analysis is dependent on the selection of the path-loss model. Although the majority of CF mMIMO works such as [12], [14] have considered another path-loss model, herein, for the sake of clarity and simplicity, we have considered a famous bounded path-loss model that will result in tractable expressions. Note that the three-slope path-loss model would be too complicated for the analysis. Also, both models provide similar insights regarding the parameters of the system under study in this work. The same reasons have contributed to the wide acceptance of the bounded model in many scenarios modeled in terms of stochastic geometry [31]. In addition, both types of distances, i.e., the distance between the m th AP located at \mathbf{x}_m in \mathbb{R}^2 and the typical user as well as the distances between the m th AP and the other users in $\mathcal{A} \setminus \{\mathbf{x}_m \in \mathcal{A}\}$ follow the uniform distribution and are independent. Also, similar to other works on CF mMIMO systems, e.g., [12]–[14], [28], we assume independent uncorrelated Rayleigh fading where the elements of \mathbf{g}_{mk} are independent and identically distributed (i.i.d.) $\mathcal{CN}(0, 1)$ random variables. Note that this assumption of uncorrelated channels is reasonable, since the service antennas (APs) in CF mMIMO systems are distributed over a large area and the AP antennas can be well separated. Hence, the set of scatterers is likely to be different for each AP and each user.

We consider a time-varying narrowband channel that is divided into coherence blocks, which are blocks of duration T_c in s and bandwidth B_c in Hz while the channels are fixed and frequency-flat. Each coherence block consists of $\tau_c = B_c T_c$ samples (channel uses) and we follow the standard block fading model where independent channel realizations appear in every block [8]. We employ the time-division-duplex (TDD) protocol with an UL training phase of τ_{tr} samples and two data transmission phases of τ_{d} (DL) and τ_{up} (UL) samples, respectively. Hence, we have $\tau_c = \tau_{\text{tr}} + \tau_{\text{up}} + \tau_{\text{d}}$ while the communication strategy is illustrated in Fig. 1. In this work, we focus on the UL training and DL data transmission phases. The duration of the latter can be expressed by $\tau_{\text{d}} = \xi(\tau_c - \tau_{\text{tr}})$ with $\xi \leq 1$, where ξ expresses the DL payload fraction transmission [27].

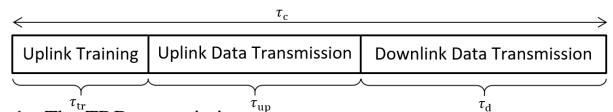


Fig. 1. The TDD transmission strategy.

III. UL CHANNEL ESTIMATION

The construction of the precoder for the DL transmission requires the channel state information, which is obtained from the UL training phase. The τ_{tr} channel uses for UL training need to be shared among all the users and there is room for τ_{tr} mutually orthogonal pilot sequences. Since $K \gg \tau_{\text{tr}}$ in most cases of interest, there will be pilot contamination. By introducing the reuse factor $\zeta = K/\tau_{\text{tr}}$, we note that ζ users share the same pilot sequences.

In the training phase of one realization of the network, the k th user transmits a normalized pilot sequence $\boldsymbol{\psi}_k \in \mathbb{C}^{\tau_{\text{tr}} \times 1}$ with $\|\boldsymbol{\psi}_k\|^2 = 1$, and the received $N \times \tau_{\text{tr}}$ channel vector by the m th AP is given by

$$\tilde{\mathbf{Y}}_m^{\text{tr}} = \sum_{i=1}^K \sqrt{\tau_{\text{tr}} \rho_{\text{tr}}} l_{mi}^{1/2} \mathbf{g}_{mi} \boldsymbol{\psi}_i^{\text{H}} + \mathbf{n}_m^{\text{tr}}, \quad (3)$$

where ρ_{tr} is the average transmit power while \mathbf{n}_m^{tr} is the $N \times \tau_{\text{tr}}$ additive noise vector at the m th AP consisted of i.i.d. $\mathcal{CN}(0, 1)$ random variables. In other words, ρ_{tr} is actually the normalized signal-to-noise ratio (SNR). By projecting $\tilde{\mathbf{y}}_{mk}^{\text{tr}}$ onto $\frac{1}{\sqrt{\tau_{\text{tr}} \rho_{\text{tr}}}} \boldsymbol{\psi}_k$, we obtain

$$\tilde{\mathbf{y}}_{mk} = \mathbf{g}_{mk} l_{mk}^{1/2} + \sum_{i \neq k} l_{mi}^{1/2} \mathbf{g}_{mi} \boldsymbol{\psi}_i^{\text{H}} \boldsymbol{\psi}_k + \frac{1}{\sqrt{\tau_{\text{tr}} \rho_{\text{tr}}}} \mathbf{n}_m^{\text{tr}} \boldsymbol{\psi}_k. \quad (4)$$

With the assumption that the channel and distances statistics are known a priori and that $\boldsymbol{\psi}_i^{\text{H}} \boldsymbol{\psi}_k \in \{0, 1\}$ for all i, k , the m th AP obtains the linear minimum mean-squared error (MMSE) estimate according to [32], i.e., $\hat{\mathbf{h}}_{mk} = \mathbb{E}[\mathbf{h}_{mk} \tilde{\mathbf{y}}_{mk}^{\text{H}}] \mathbb{E}^{-1}[\tilde{\mathbf{y}}_{mk} \tilde{\mathbf{y}}_{mk}^{\text{H}}] \tilde{\mathbf{y}}_{mk}$. Thus, we have

$$\hat{\mathbf{h}}_{mk} = \frac{l_{mk}}{\sum_{i=1}^K |\boldsymbol{\psi}_i \boldsymbol{\psi}_k^{\text{H}}|^2 l_{mi} + \frac{1}{\tau_{\text{tr}} \rho_{\text{tr}}}} \tilde{\mathbf{y}}_{mk}. \quad (5)$$

The estimation error vector $\tilde{\mathbf{e}}_{mk} = \mathbf{h}_{mk} - \hat{\mathbf{h}}_{mk}$ is independent of $\hat{\mathbf{h}}_{mk}$. Moreover, it follows that $\mathbf{h}_{mk} \in \mathbb{C}^{N \times 1} \sim \mathcal{CN}(\mathbf{0}, l_{mk} \mathbf{I}_N)$, $\hat{\mathbf{h}}_{mk} \in \mathbb{C}^{N \times 1} \sim \mathcal{CN}(\mathbf{0}, \sigma_{mk}^2 \mathbf{I}_N)$ and $\tilde{\mathbf{e}}_{mk} \in \mathbb{C}^{N \times 1} \sim \mathcal{CN}(\mathbf{0}, \tilde{\sigma}_{mk}^2 \mathbf{I}_N)$, where $\sigma_{mk}^2 = \frac{l_{mk}^2}{d_m}$ and $\tilde{\sigma}_{mk}^2 = l_{mk} \left(1 - \frac{l_{mk}}{d_m}\right)$ with $d_m = \left(\sum_{i=1}^K |\boldsymbol{\psi}_i \boldsymbol{\psi}_k^{\text{H}}|^2 l_{mi} + \frac{1}{\tau_{\text{tr}} \rho_{\text{tr}}}\right)$.

IV. DL TRANSMISSION

We now consider the DL transmission in one realization of the network, where the APs have multiple antennas and are PPP distributed. The goal is to derive the achievable SE with conjugate beamforming, taking into account the effect of pilot contamination as well as the spatial randomness of the APs. "The choice of conjugate beamforming relies on its indication for distributed architectures due to no need for CSI exchange among the APs and the central unit [12]. Also, conjugate beamforming performs well in both CF mMIMO systems and SCs, and it provides the derivation of closed-form tractable expressions⁷.

⁷Notably, the impact of zero-forcing (zero-forcing) in terms of analytical closed-form results is the topic of ongoing research. Therein, the more robust regularized ZF, which is indicated for better performance, is also studied but in terms of Monte-Carlo (MC) simulations since it does not provide closed-form expressions.

Note that although the choice of zero-forcing (zero-forcing) or regularized ZF is indicated for better performance, their applications would not allow the derivation of any tractable closed-form expressions, which is one of the main contributions of this work.

The received signal by the typical user is given by

$$y_k^{\text{d}} = \sqrt{\rho_{\text{d}}} \sum_{i \in \Phi_{\text{AP}}} \tilde{\mathbf{h}}_i^{\text{H}} \mathbf{s}_i + z_k^{\text{d}} \quad (6)$$

$$= \sqrt{\rho_{\text{d}}} \sum_{m=1}^M \mathbf{h}_{mk}^{\text{H}} \mathbf{s}_m + z_k^{\text{d}}. \quad (7)$$

In (6), the vector $\tilde{\mathbf{h}}_i$ describes the channel between the i th AP located at $\mathbf{x}_i \in \mathbb{R}^2$ and the typical user including small-scale fading and path-loss, $\rho_{\text{d}} > 0$ denotes the corresponding transmit power, while \mathbf{s}_i is the transmitted signal from the i th AP, and $z_k^{\text{d}} \sim \mathcal{CN}(0, 1)$ is the additive white Gaussian noise at the k th user. Since a realization of the system includes M APs, the signal model described by (6) can be written as in (7). Notably, the number M is a random variable changing in every spatial realization of the APs. In (7), the vector \mathbf{h}_{mk} expresses the channel between the m th AP and the typical user while \mathbf{s}_m is the transmit signal from the m th AP, which is written as

$$\mathbf{s}_m = \sum_{k=1}^K \sqrt{\eta_{mk}} \mathbf{f}_{mk} q_k, \quad (8)$$

where $q_k \in \mathbb{C}$ is the normalized transmit data symbol for user k satisfying $\mathbb{E}[|q_k|^2] = 1$. The vector $\mathbf{f}_{mk} = \hat{\mathbf{h}}_{mk} \in \mathbb{C}^N$ expresses the linear precoder. Also, we denote $\eta_{mk} = \mu \sigma_{mk}^{-4}$ with μ obtained by means of the constraint of the transmit power $\mathbb{E}\left[\frac{\rho_{\text{d}}}{K} \mathbf{s}_m \mathbf{s}_m^{\text{H}}\right] = \rho_{\text{d}}$. This selection regarding η_{mk} aims at easing the following algebraic manipulations. Actually, it corresponds to a statistical channel inversion power-control policy [25]. It allows each AP to allocate more power to the most distant users and less power to the closest ones. Note that the scaling does not result in any loss in the performance since the parameter μ is changed accordingly.

Henceforth, for the sake of algebraic manipulations, we denote $\mathbf{h}_k = [\mathbf{h}_{1k}^T \cdots \mathbf{h}_{Mk}^T] \sim \mathcal{CN}(\mathbf{0}, \mathbf{L}_k)$, $\hat{\mathbf{h}}_k = [\hat{\mathbf{h}}_{1k}^T \cdots \hat{\mathbf{h}}_{Mk}^T] \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{\Phi}_k)$ and $\tilde{\mathbf{e}}_k \in \mathbb{C}^{\mathcal{W} \times 1} \sim \mathcal{CN}(\mathbf{0}, \mathbf{L}_k - \boldsymbol{\Phi}_k)$. The matrices $\mathbf{L}_k \in \mathbb{C}^{\mathcal{W} \times \mathcal{W}}$, $\boldsymbol{\Phi}_k = \mathbf{L}_k^2 \mathbf{D}^{-1} \in \mathbb{C}^{\mathcal{W} \times \mathcal{W}}$, and $\mathbf{D} \in \mathbb{C}^{\mathcal{W} \times \mathcal{W}}$ are block diagonal matrices with elements given by the matrices $[\mathbf{L}_k]_{ww} = l_{mk} \mathbf{I}_N$, $[\boldsymbol{\Phi}_k]_{ww} = \sigma_{mk}^2 \mathbf{I}_N$, $[\mathbf{D}]_{ww} = d_m \mathbf{I}_N$, and $[\mathbf{D}]_{ww} = d_m \mathbf{I}_N$, respectively, for $w = 1, \dots, \mathcal{W}$ and $\mathcal{W} = MN$. We also define $\mathbf{C}_k = \boldsymbol{\Phi}_k^{-1}$ with $[\mathbf{C}_k]_{ww} = c_{mk} \mathbf{I}_N$, where $c_{mk} = \sigma_{mk}^{-2}$.

After substituting (8) into (7), the received signal by the typical user is given by

$$y_k^{\text{d}} = \sqrt{\rho_{\text{d}}} \left(\mathbb{E} \left[\sum_{m=1}^M \eta_{mk}^{1/2} \mathbf{h}_{mk}^{\text{H}} \hat{\mathbf{h}}_{mk} \right] q_k + \sum_{m=1}^M \eta_{mk}^{1/2} \mathbf{h}_{mk}^{\text{H}} \hat{\mathbf{h}}_{mk} q_k \right. \\ \left. - \mathbb{E} \left[\sum_{m=1}^M \eta_{mk}^{1/2} \mathbf{h}_{mk}^{\text{H}} \hat{\mathbf{h}}_{mk} \right] q_k + \sum_{i \neq k} \sum_{m=1}^M \eta_{mi}^{1/2} \mathbf{h}_{mk}^{\text{H}} \hat{\mathbf{h}}_{mi} q_i \right) + z_k^{\text{d}}, \quad (9)$$

where we have written (9) similar to [33], in order to derive the SINR based on the fact that the users do not have any

knowledge of the instantaneous CSI, but they are aware of its statistics⁸. Note that the second term in (9) expresses the desired signal while the fourth term describes the multi-user interference. By applying the well-established bounding technique in [33], we consider that (9) represents a single-input single-output (SISO) system, where the APs treat the unknown terms as uncorrelated additive noise. Thus, we obtain the effective SINR of the DL transmission from all the multi-antenna APs to the typical user, conditioned on the number of APs and their distances from the users, as

$$\tilde{\gamma}_k = \frac{\left| \mathbb{E} \left[\mathbf{h}_k^H \mathbf{C}_k \hat{\mathbf{h}}_k \right] \right|^2}{\sum_{i=1}^K \mathbb{E} \left[\left| \mathbf{h}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] - \left| \mathbb{E} \left[\mathbf{h}_k^H \mathbf{C}_k \hat{\mathbf{h}}_k \right] \right|^2 + \frac{1}{\mu p_d}}. \quad (10)$$

Notably, the matrices in (10) are random because they include the number of APs and the distances between the APs and the users, being random variables changing in each realization.

Proposition 1: Given a realization of the network with M APs and K users, the effective SINR of the DL transmission at the typical user in a CF mMIMO system, accounting for pilot contamination and conjugate beamforming, is given by (11).

Proof: See Appendix A. ■

Remark 1: The scaling in the numerator with N corresponds to the array gain resulting from the coherent transmission of the N antennas per AP. Moreover, the summations in the denominator are over the number of users K because as their number increases, the interference increases.

V. EE ANALYSIS

In this section, we provide the definition of the EE of CF mMIMO systems where the APs locations follow a PPP distribution. Specifically, we focus on the analytical derivation of the DL EE by first obtaining a lower bound on the average ASE, and then, presenting a realistic power consumption model. The power consumption is expected to increase rapidly with the number of APs, i.e., their density. Hence, it is of paramount importance to quantify the relevant efficiency of a CF mMIMO system.

Definition 1: The EE, denoted by \bar{E} , expresses the amount of reliably transmitted information per unit of energy, which is defined mathematically as

$$\bar{E} [\text{bit/Joule}] = \frac{B_w [\text{Hz}] \cdot \text{ASE} \left[\text{bit/s} / (\text{Hz} \cdot \text{km}^2) \right]}{\text{APC} \left[\text{W/km}^2 \right]}, \quad (12)$$

where B_w , ASE, and APC describe the transmission bandwidth, the ASE, and the area power consumption (APC), respectively.

We continue with the derivations of ASE and APC.

⁸Although, in general, channel hardening does not appear in CF mMIMO systems with single-antenna APs according to [13] and (9), we exploit its property because the proposed model considers multi-antenna APs. In fact, in [13], it was shown that if $N \geq 5$, channel hardening is met. Taking into account for this limitation, numerical results, provided in Section VII, reveal that values of interest regarding N is more than 10, which corroborates our analysis taking advantage of channel hardening.

A. Area Spectral Efficiency

Taking advantage of the property of the typical user, stating that it is statistically equivalent with any other user in the network, the ASE is provided by

$$\text{ASE} = K' \bar{R} \left[\text{bit/s} / (\text{Hz} \cdot \text{km}^2) \right], \quad (13)$$

where $K' = K/S$ is the number of users per area S , and \bar{R} is the average SE per user. We have $\bar{R} = \bar{R}_k$, where \bar{R}_k , provided below, is the average DL SE of user k , being statistical equivalent with any other user in the network.

Remark 2: Contrary to the common definition for the ASE in cellular systems with no cooperation [25], [27], the CF mMIMO architecture necessitates a new definition. Specifically, in CF mMIMO systems each user receives joint transmission from multiple sources (APs), and the received SINR at the user is obtained from the sum of received signals from all these serving APs. Therefore, this received SINR is not the same as the received SINR computed in a single BS association network. Consequently, the definition of ASE [25], where the received user rate (i.e., per transmission link rate) is multiplied with the AP density does not hold in this scenario.

Since the DL capacity for this network including imperfect CSI is not known, we follow the common approach, especially in the area of mMIMO [6], [34], focusing on the derivation of achievable lower bounds on the ergodic capacity. In particular, the following lemma provides a tractable lower bound on the ergodic capacity for any given realization of Φ_{AP} .

Lemma 1 ([35]): A lower bound on the DL ergodic channel capacity of the typical user k in a CF mMIMO system with conjugate beamforming and PPP distributed APs for any given realization of Φ_{AP} is provided by

$$\bar{R}_k = \left(1 - \frac{K}{\zeta \tau_c} \right) \log_2 (1 + \tilde{\gamma}_k) \quad [\text{b/s/Hz}], \quad (14)$$

where K is the number of users, ζ is the pilot reuse factor, and τ_c is the channel coherence interval in number of samples while $\tilde{\gamma}_k$ is given by (11).

The average SE per user is obtained by applying the expectation at (14) over the APs locations. We resort to Jensen's inequality to derive a closed-form lower bound for the DL achievable R_k and avoid intractable lengthy numerical integral evaluations with respect to the APs distances.

Theorem 1: A lower bound on the DL average SE per user with conjugate beamforming precoding in a CF mMIMO system with multi-antenna APs is obtained by

$$R_k = \left(1 - \frac{K}{\zeta \tau_c} \right) \log_2 (1 + \gamma_k) \quad \text{b/s/Hz}, \quad (15)$$

where $\gamma_k = 1/\tilde{\gamma}_k$ with $\tilde{\gamma}_k$ given by

$$\tilde{\gamma}_k = \sum_{j=1}^K |\psi_j \psi_k^H|^2 \left(\frac{\alpha - 2}{\alpha \pi N p_d} + K - 1 \right) + \frac{\zeta}{\alpha \pi K \rho_{\text{tr}}} \left((K - 1)(\alpha - 2) + \frac{(\alpha - 1)}{N p_d} \right) + \lambda_{\text{AP}} (K - 1). \quad (16)$$

Proof: See Appendix B. ■

Notably, if we shed further light into (15), we observe that the sum SE is a strictly quasi-concave function of the number

$$\bar{\gamma}_k = \frac{M^2 N}{\sum_{i=1}^K \text{tr} \left(\mathbf{C}_i \left(N \mathbf{L}_k + \frac{1}{K p_d} \mathbf{I}_M \right) \right) + N \sum_{i \neq k}^K \text{tr}^2 \left(\mathbf{L}_k \mathbf{L}_i^{-1} \right) - N \text{tr} \left(\mathbf{D} \mathbf{L}_k^{-1} \right) + M}. \quad (11)$$

of users K while the optimal number of antennas per AP depends on the AP density and the quality of CSI in terms of N and ζ , respectively. These observations are in line with [1], accounting also for the spatial AP randomness.

Although we have applied the law of large numbers regarding the number of APs during the derivation of this proof, it is known that this law is applicable and valid in the case of a finite number of APs obeying to $M > 8$ [36]. Obviously, this range is of practical interest in CF mMIMO systems. The agreement of the analytical results with MC simulations in Section VII for finite system dimensions confirms this assertion. Thus, Theorem 1 and the following results describe realistic systems of finite dimensions.

B. Area Power Consumption

The sources of the area power consumption of a CF mMIMO system are the power usage during the transmission P_{TX} and the circuitry of the system P_{CPC} . Following a similar approach to [27], [37] but specialized to CF massive systems, we have

$$\text{APC} = \lambda_{\text{AP}} \left(\frac{1}{\alpha_{\text{eff}}} P_{\text{TX}} + P_{\text{CPC}} \right), \quad (17)$$

where $\alpha_{\text{eff}} \in (0, 1]$ is the power amplifier efficiency. Note that P_{TX} concerns both the average powers for the UL pilot and DL payload transmissions. Regarding P_{CPC} , it describes the circuitry dissipation in terms of cooling, power supply, backhaul signaling, digital signal processing, etc. Although the majority of works assume that P_{CPC} is a fixed constant, this is not a realistic assumption, and obviously, not a good design methodology. In practice, each antenna is accompanied by dedicated circuits that contribute to the system power consumption. Above this, if APC was independent of N , the ASE, increasing with N , would result in an unbounded EE as N increases, which is irrational [37]. Hence, it is of dire necessity to incorporate in our EE analysis an accurate model for the power consumption.

Proposition 2: A realistic model for the DL APC of CF mMIMO systems is given by

$$\text{APC}(\boldsymbol{\theta}) = \lambda_{\text{AP}} (C_0 + C_1 K + C_2 K^2 + D_0 N + D_1 N K - D_2 N K^2) + \mathcal{B} B_w \text{ASE}, \quad (18)$$

where $C_0 = P_{\text{FP}} + P_{\text{LO}}$, $C_1 = \frac{B_w}{7 L_{\text{AP}} \tau_c} - \frac{\xi \rho_d}{\alpha_{\text{eff}} \zeta \tau_c} + P_{\text{UE}}$, $C_2 = \frac{1}{\alpha_{\text{eff}} \zeta \rho_{\text{tr}} \tau_c}$, $D_0 = P_{\text{AP}}$, $D_1 = \frac{3 B_w}{L_{\text{AP}}} + \frac{3 B_w}{L_{\text{AP}} \tau_c}$, $D_2 = \frac{3 B_w (\xi - 1)}{L_{\text{AP}} \zeta \tau_c}$, and $\mathcal{B} = (P_{\text{COD}} + P_{\text{DEC}} + P_{\text{BT}})$.

Proof: See Appendix C. ■

It is worthwhile to mention that (18) is written in a polynomial structure that will facilitate the optimization taking place in the following section.

VI. EE MAXIMIZATION

This section elaborates on the main objective of this work, which is the maximization of the constrained EE with respect to the parameters defining the size of the network (e.g., the AP density and the number of users)⁹. In other words, we scrutinize the tuple of system parameters $\boldsymbol{\theta} = (\zeta, \lambda_{\text{AP}}, K, N)$ obeying to the problem

$$\boldsymbol{\theta}^* = \arg \max_{\boldsymbol{\theta} \in \Theta} \text{EE}(\boldsymbol{\theta}) = \frac{B_w \text{ASE}(\boldsymbol{\theta})}{\text{APC}(\boldsymbol{\theta})} \quad (19)$$

subject to $\bar{\gamma}_k(\boldsymbol{\theta}) = \gamma_0$,

where $\text{ASE}(\boldsymbol{\theta}) = K' R$ with $R = R_k$, where R_k is given by Theorem 1, $\text{APC}(\boldsymbol{\theta})$ is provided by Proposition 2, $\bar{\gamma}_k$ is obtained by Theorem 1 while $\gamma_0 > 0$ is a design parameter. The set Θ , including the feasible parameters values, is defined as $\Theta = \{\boldsymbol{\theta} : \lambda_{\text{AP}} \geq 0, \zeta \geq 1, K/\zeta \leq \tau_c, (K, N) \in \mathbb{Z}_+\}$. The constraint in (19) prevents from an optimal tuple of parameters with a low unacceptable achievable rate while it demands a specific QoS [25], [27].

We aim at solving (19) for either $\lambda_{\text{AP}}, N, K$ when the remaining parameters are fixed. The advantage of this approach is to obtain closed-form expressions for the optimal EE and to shed light into the interplay among these parameters.

A. Feasibility

The optimization problem in (19) is feasible for a certain range of values of γ_0 because of the multiuser interference.

Lemma 2: The feasibility range of values of γ_0 , used in the maximization problem for CF mMIMO systems (19), can be described by

$$\gamma_0 < \frac{1}{\lambda_{\text{AP}}}. \quad (20)$$

Proof: In order to obtain the range of values of γ_0 , we simplify the expression of the SINR, being the inverse of (16) by noticing that it is a monotonically increasing function of N . Hence, deriving its upper limit as $N \rightarrow \infty$ ¹⁰, we obtain (21). Since the upper limit is a decreasing function of the optimizable variable ζ , we exploit the constraint $\zeta = K/\tau_{\text{tr}}$ by taking its minimal value when $K = 1$, and we obtain the feasible γ_0 . ■

This lemma reveals that the upper limit of the SINR depends only on λ_{AP} as $N \rightarrow \infty$. In the case of CF mMIMO, the typical value concerning the number of APs is 100 – 200 [12] which is equivalent to a density $\lambda_{\text{AP}} \approx 10^{-4} \text{ m}^{-2}$. In such

⁹The study of the EE optimization with regard to the transmit power in CF mMIMO systems with PPP distributed APs is the topic of our ongoing research.

¹⁰Although the assumption of an infinite number of antennas per AP, i.e., $N \rightarrow \infty$ is impractical, it is used here just for showing the feasibility range of the target SINR and has no impact on the main results since the following analysis is for finite N .

$$\lim_{N \rightarrow \infty} \bar{\gamma}_k = \frac{\alpha \pi \rho_{\text{tr}} K}{\alpha \pi K \left(\sum_{j=1}^K |\psi_j \psi_k^{\text{H}}|^2 (K-1) + K \lambda_{\text{AP}} \right) \rho_{\text{tr}} + (\alpha - 2) (K-1) \zeta}. \quad (21)$$

case, e.g., $\lambda_{\text{AP}} = 10^{-4} \text{ m}^{-2}$, the average SE per user is $\log_2(1 + 100) \approx 13.29$ b/s/Hz. This value, showing the feasibility of the optimization problem described by (19), is larger than the SE of currently applied systems [38]. Hence, the optimization problem under study is quite meaningful for practical systems.

B. Optimal Pilot Reuse Factor

Herein, we derive the optimal pilot reuse factor ζ^* while the rest of the parameters are fixed.

Theorem 2: Let any set of $\{\lambda_{\text{AP}}, K, N\}$ resulting in the feasibility of the maximization of EE given by (19). The optimal pilot reuse factor, satisfying the SINR constraint, is obtained by

$$\zeta^* = \frac{\alpha \pi K N \rho_{\text{tr}} \rho_{\text{d}} - \gamma_0 Q_1}{\gamma_0 Q_2}. \quad (22)$$

Proof: The reuse factor ζ^* is obtained by focusing on the constraint and collecting the terms including ζ in the SINR given by $\bar{\gamma}_k = 1/\check{\gamma}_k$ as

$$\gamma_0 = \frac{\alpha \pi N \rho_{\text{tr}} \rho_{\text{d}}}{Q_1 - \zeta Q_2}, \quad (23)$$

where we set

$$\begin{aligned} Q_1 &= K \rho_{\text{tr}} \left((\alpha - 2) \sum_{j=1}^K |\psi_j \psi_k^{\text{H}}|^2 \right. \\ &\quad \left. + \alpha \pi N \rho_{\text{d}} (K - 1) \left(\sum_{j=1}^K |\psi_j \psi_k^{\text{H}}|^2 + \lambda_{\text{AP}} \right) \right), \\ Q_2 &= (\alpha - 1 + N \rho_{\text{d}} (\alpha - 2) (K - 1)) / K, \end{aligned} \quad (24)$$

and we solve (23) with respect to ζ . ■

Theorem 2 provides the dependence of ζ^* on the rest of the system parameters. According to its physical interpretation, a smaller pilot reuse factor, meaning a larger training phase, results in both more precise channel estimation and less pilot contamination. Intuitively, a better channel estimation increases the SE, or equivalently, a better SINR constraint γ_0 is allowed, which comes to agreement with (22). It is shown that ζ^* is a decreasing function of Q_1 and Q_2 , which both are increasing functions of K . However, a larger K means higher interference, requiring a better channel estimation, i.e., a lower ζ which admits to the dependence shown by (22).

C. Optimal APs Density

After plugging (22) into the optimization problem, (19) is written as

$$\begin{aligned} \text{EE}(\zeta^*, K, N) &= \frac{B_w \text{ASE}(\zeta^*, K, N)}{\text{APC}(\zeta^*, K, N)} \\ \text{subject to } 1 &\leq \frac{\alpha \pi N \rho_{\text{tr}} \rho_{\text{d}} - \gamma_0 Q_1}{\gamma_0 Q_2} \leq \frac{K}{\tau_c}. \end{aligned} \quad (25)$$

Theorem 3: Let any set of $\{K, N\}$ keeping the optimization problem (25) feasible. For fixed K and N , the EE is maximized by

$$\lambda_{\text{AP}}^* = \min(\max(\lambda_{\text{AP}0}, \lambda_{\text{AP}1}), \lambda_{\text{AP}2}), \quad (26)$$

where

$$\lambda_{\text{AP}0} = \frac{(a_1 + a_3) G + \sqrt{a_2 a_3 a_4 (a_1 + a_3) G}}{a_2 a_4 G} \quad (27)$$

with

$$G = a_2 (a_4 + a_5 + a_6 K \log(1 + \gamma_0)) \quad (28)$$

while $\lambda_{\text{AP}1} = \frac{a_3 - a_1}{a_2}$, $\lambda_{\text{AP}2} = \frac{\tau_c / (K a_3) + a_1}{a_2}$, and the parameters $\{a_i\}$ are provided in Table I.

Proof: Both ASE and APC include the term $\zeta^* \tau_c / K$. Hence, we proceed with its computation which gives $\zeta^* \tau_c / K = \frac{a_2 \lambda_{\text{AP}} - a_1}{a_3}$. Then, after substituting this term into the objective function of (25), the EE becomes

$$\text{EE}(\zeta^*) = \frac{\frac{K \xi}{S \lambda_{\text{AP}}} (1 - \bar{a}) \log_2(1 + \gamma)}{a_4 + a_5 \bar{a} + a_6 K \xi (1 - \bar{a}) \log_2(1 + \gamma)}, \quad (29)$$

where $\bar{a} = \frac{a_3}{a_2 \lambda_{\text{AP}} - a_1}$. Following the approach in [37, Lem. 3], it can be shown that (29) is a quasi-concave function of λ_{AP} . Thus, (27) is obtained by taking the first derivative of (29) and equating to zero. Given that the constraint in (25) depends on λ_{AP} , we obtain $\lambda_{\text{AP}1}$ and $\lambda_{\text{AP}2}$. ■

D. Optimal Number of AP Antennas and Users

The optimal values of N and K are found by means of the maximization problem (25) in the case of optimal ζ^* . Initially, we consider the integer-relaxed problem where K and N can be any positive scalars, but then, we select the corresponding integer values.

Theorem 4: Let the maximization problem (19) with λ_{AP} , K , and N real variables. For any fixed λ_{AP} , $K > 0$, the optimal number of AP antennas N^* is given by

$$N^* = \min(\max(N_0, N_1), N_2) \quad (30)$$

with $N_0 = \frac{q_1 - \sqrt{q_3}}{q_2}$ while $N_1 = \frac{b_1 + b_3}{b_4 - b_2}$ and $N_2 = \frac{\frac{K}{\tau_c} b_1 + b_3}{b_4 - \frac{K}{\tau_c} b_2}$, where $q_1 = b_3^2 \bar{d}_{11} + b_1 K (b_1 d_{22} K^2 + b_4 \bar{c}_{11}) + b_3 K b_2 \bar{c}_{11} + b_1 (\bar{d}_{11} + d_{22} K)$, $q_2 = 2(b_3 + b_1 K)(b_4 - b_2 K)(b_2 d_{22} K^2 - b_4 \bar{d}_{11})$, $q_3 = q_1^2 + 4q_2(b_2 d_{22} K^2 - b_4 \bar{d}_{11})$ with $\bar{c}_{11} = c_{12} - (c_{11} + c_{22}) K$, $c_{11} = \left(\frac{B_w}{7 L_{\text{AP}} \tau_c} + P_{\text{UE}} \right) / \lambda_{\text{AP}}$, $c_{12} = \left(\rho_{\text{d}} \frac{1 - \xi}{\alpha_{\text{eff}}} - C_0 \right) / \lambda_{\text{AP}}$, $c_{22} = \frac{\rho_{\text{tr}}}{\alpha_{\text{eff}} \lambda_{\text{AP}}}$, $d_{22} = \frac{3 B_w \xi}{L_{\text{AP}} \lambda_{\text{AP}}}$, $\bar{d}_{11} = (D_0 + D_1 K) / \lambda_{\text{AP}}$, and the parameters $\{b_i\}$ are provided in Table I.

Proof: Similar to the proof of Theorem 3, we have $\zeta^* \tau_c / K = \frac{b_1 + b_2 N}{b_4 N - b_3}$, where the parameters $\{b_i\}$ are provided

TABLE I
OPTIMIZATION PARAMETERS FOR OPTIMAL APS DENSITY λ_{AP} AND NUMBER OF ANTENNAS N

| Parameter | Value | Parameter | Value |
|-----------|---|-----------|---|
| a_1 | $\rho_{\text{tr}}\tau_{\text{c}}K \left(\alpha\pi N\rho_{\text{d}} - ((\alpha-2) + \alpha\pi N\rho_{\text{d}})\gamma \sum_{j=1}^K \psi_j\psi_k^{\text{H}} ^2 \right)$ | b_1 | $K(\alpha-1)$ |
| a_2 | $\alpha\pi\gamma\rho_{\text{d}}\rho_{\text{tr}}\tau_{\text{c}}KN(K-1)$ | b_2 | $\rho_{\text{d}}K(K-1)(\alpha-2)$ |
| a_3 | $\gamma KN\rho_{\text{d}}(\alpha-1 + (K-1)(\alpha-2))$ | b_3 | $\alpha\pi\rho_{\text{d}}\rho_{\text{tr}}\tau_{\text{c}} \left(1 - \gamma \left(\sum_{j=1}^K \psi_j\psi_k^{\text{H}} ^2 + \lambda_{\text{AP}} \right) (K-1) \right)$ |
| a_4 | $C_0 + C_{11}K + N(D_0 + D_1K)$ | b_4 | $\rho_{\text{tr}}\tau_{\text{c}}\gamma(\alpha-2) \sum_{j=1}^K \psi_j\psi_k^{\text{H}} ^2$ |
| a_5 | $\frac{1}{\alpha_{\text{eff}}} ((\xi-1)\rho_{\text{d}} + K\rho_{\text{tr}}) - \frac{3BK\xi}{L_{\text{AP}}}$ | b_5 | $(C_0 + C_{11}K)/\lambda_{\text{AP}}$ |
| a_6 | $P_{\text{COD}} + P_{\text{DEC}} + P_{\text{BT}}$ | b_6 | $(D_0 + D_1K - D_2K^2)/\lambda_{\text{AP}}$ |
| a_7 | — | b_7 | $(c_{22} - d_{22} - c_{12})K^2/\lambda_{\text{AP}}$ |
| a_8 | — | b_8 | $(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{BT}})/\lambda_{\text{AP}}$ |

TABLE II
OPTIMIZATION PARAMETERS FOR OPTIMAL NUMBER OF USERS K

| Parameter | Value |
|-----------|--|
| e_1 | $\gamma N\rho_{\text{d}}(\alpha-2)$ |
| e_2 | $\gamma N\rho_{\text{d}}(1-3(\alpha-2))$ |
| e_3 | $2\gamma(1-\alpha) - 2b_1$ |
| e_4 | $-\alpha\pi N\rho_{\text{tr}}\tau_{\text{c}}\rho_{\text{d}}\gamma(\tau_{\text{tr}} + \lambda_{\text{AP}} - 1)$ |
| e_5 | $\gamma\rho_{\text{tr}}\tau_{\text{c}}((\alpha-2)(\tau_{\text{tr}}+1) + \alpha\pi N\rho_{\text{tr}}\tau_{\text{c}}\rho_{\text{d}}(\tau_{\text{tr}}+\gamma)) - e_6$ |
| e_6 | $(\alpha-2)(\gamma-3\tau_{\text{tr}})$ |
| e_7 | $(C_0 + D_0N)/\lambda_{\text{AP}}$ |
| e_8 | $(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{BT}})/\lambda_{\text{AP}}$ |
| e_9 | $(C_2 - D_2N)/\lambda_{\text{AP}}$ |
| e_{10} | $(P_{\text{COD}} + P_{\text{DEC}} + P_{\text{BT}})/\lambda_{\text{AP}}$ |

in Table I. Then, after substituting this term into the objective function of (25), the EE becomes

$$\text{EE}(\zeta^*) = \frac{\frac{K\xi}{S}(1-\bar{b})\log_2(1+\gamma)}{b_5 + b_6N + b_7\bar{b} + b_8K\xi(1-\bar{b})\log_2(1+\gamma)}, \quad (31)$$

which represents a quasi-concave function of N . Note that $\bar{b} = \frac{b_1 + b_2N}{b_4N - b_3}$. The optimal value of N is obtained by computing its first derivative with respect to N and equating it to zero. The resultant value, satisfying the unconstraint problem, is given by (30). Taking into account for the constraint in (25), this can be written as $\frac{\tau_{\text{c}}}{K} \leq \frac{b_1 + b_2N}{b_4N - b_3} \leq 1$, which results in $N_1 \leq N^* \leq N_2$. ■

Theorem 5: Let the maximization problem (19) with λ_{AP} , K and N real variables. For any fixed λ_{AP} , $N > 0$, the optimal number of users K^* is given by

$$K^* = \max(K_2, \max(K_{1,1}, \min(K_0, K_{1,2}))), \quad (32)$$

where K_0 is one of the real roots of a quintic equation, i.e., a polynomial of degree five given by $\sum_{i=0}^5 p_i x_i = 0$ with $p_0 = e_1 - e_2$, $p_1 = (e_1 - e_2) + 2e_3$, $p_2 = 3(e_3 - 2e_1)$, $p_3 = (e_2 + e_1) - e_1e_3$, $p_4 = 2e_4(e_1 - e_2)$, $p_5 = (e_2 - e_1) - e_1$. Also, we have $K_2 = \frac{e_4 \frac{K}{\tau_{\text{c}}} - e_2}{e_1 - e_3 \frac{K}{\tau_{\text{c}}}}$ and

$$K_{1,1} = \frac{-(e_1 - e_2) - \sqrt{(e_1 - e_2)^2 - 4e_3e_4}}{2e_4} \quad (33)$$

$$K_{1,2} = \frac{-(e_1 - e_2) + \sqrt{(e_1 - e_2)^2 - 4e_3e_4}}{2e_4}. \quad (34)$$

Proof: We notice that the term $A = \sum_{i=1}^K |\psi_i^{\text{H}}\psi_k|^2$, appearing in $\zeta^*\tau_{\text{c}}/K$, depends on K by means of its superscript. In fact, $\zeta^*\tau_{\text{c}}/K$ is an increasing function regarding A . Hence,

we apply the bound on A by using the Welch inequality [39], and we obtain

$$A \geq \frac{\tau_{\text{tr}}(K-3) + K-1}{\tau_{\text{tr}}(K-2)} \quad (35)$$

since the summation becomes $\sum_{i \neq k}^K |\psi_i^{\text{H}}\psi_k|^2 = \frac{K-1-\tau_{\text{tr}}}{\tau_{\text{tr}}(K-2)}$ by using the inequality. Substituting (35) into (31) and rearranging with respect to K , the objective function can be written as

$$\text{EE}(\zeta^*) = \frac{\frac{K\xi}{S}(1-\bar{e})\log_2(1+\gamma)}{e_7 + e_8K + e_9K^2 + e_{10}\xi(1-\bar{e})\log_2(1+\gamma)}, \quad (36)$$

where $\bar{e} = \frac{e_1K^2 + e_2K + e_3}{e_4K^2 + e_5K + e_6}$ while the parameters $\{e_i\}$ are provided in Table II. Taking the first derivative of (36) with respect to K and equating it to zero, we obtain a polynomial fifth degree with roots provided by an exhaustive search over the domain set while using a bisection method and the help of Mathematica [40]. We obtain three real roots and one pair of complex roots. Note that the constraint results in K_2 . ■

VII. NUMERICAL RESULTS

This section presents illustrations of the analytical results provided by means of Theorems 2-5 concerning the optimal EE. Notably, the tightness of the derived bounds, denoting their values as good approximations, is demonstrated in Fig. 4 by MC simulations. For the sake of comparison, we have considered a conventional ‘‘cellular’’ mMIMO scenario and a SCs architecture. Especially, in SCs, the effective channel power does not harden while in the mMIMO architectures we observe the signal power tending to its mean as the number of APs becomes large [6], [12]. Hence, SCs require both UL and DL training phases, i.e., the length for training in SCs is doubled. Furthermore, CF mMIMO systems enjoy favorable propagation, and thus, they can achieve optimal performance with simple linear processing. Also, the co-processing, taking place in CF systems, suppresses the inter-cell interference degrading the performance of SCs [21].

We consider a sufficiently large squared area of $S = 1 \text{ km}^2$, where the locations of the APs are simulated as realizations of the PPP Φ_{AP} with density $\lambda_{\text{AP}} = 100 \text{ APs/km}^2$ based on a wraparound topology to keep the translation invariance. We assume that the system bandwidth is $B_{\text{w}} = 20 \text{ MHz}$ and that each coherence block consists of $\tau_{\text{c}} = 200$ samples corresponding to a coherence bandwidth of 200 KHz and a coherence time of 1 ms [12]. Moreover, we assume that $N = 20$ antennas per AP and $K = 10$ users in total while $\zeta = 4$. Also, we assume that $\rho_{\text{tr}} = 100 \text{ mW}$, $\rho_{\text{d}} = 200 \text{ mW}$,

$\alpha = 4$, and $\xi = 1/3$. Moreover, the normalized UL training transmit power per pilot symbol $\bar{\rho}_{\text{tr}}$ and DL transmit power $\bar{\rho}_{\text{d}}$ result by dividing ρ_{tr} and p_{d} with the noise power N_{P} given in W by $N_{\text{P}} = \kappa_{\text{B}} W_{\text{c}} T_0 N_{\text{F}}$. For the sake of reference, the descriptions and values of the various system parameters are found in Table III unless otherwise stated. Note that the circuit power parameters have been taken from [8].

Firstly, we assess the EE by varying the pilot reuse factor ζ and AP density λ_{AP} for a given pair of K , N in a CF mMIMO setting. Specifically, in Fig. 2 and in line with Theorem 2, it is shown that the EE is a pseudo-concave function with respect to ζ with a unique global maximum at $\zeta^* = 3$ while the corresponding optimal EE is $\text{EE}^* = 5.92$ Mbit/Joule. Regarding the AP density, the EE is a quasi-concave function with respect to λ_{AP} as was stated by Theorem 3. Nevertheless, this figure shows the optimal AP density to achieve maximum EE. Hence, we observe that when $\lambda_{\text{AP}} = 25$ APs/km², the EE takes its maximum value. It is worthwhile to mention that the optimal λ_{AP} depends on fundamental system parameters such as the transmit power and the number of antennas per AP as the corresponding theorem shows.

In Fig. 3(a), we illustrate the cellular scenario, where we have assumed an AP with $N = 20$ antennas is located per cell and $K = 10$ users are served in total. In fact, we have relied on a similar work [25], studying the UL transmission of a cellular network with PPP distributed BSs, in order to simulate the EE for the DL. Notably, the outperformance of the CF mMIMO setting is depicted. In particular, in CF mMIMO systems, the EE is higher and the required AP density is much lower. In addition, in the case of SCs, studied in Fig. 3(b), we have considered the system model in [41], where independent users are associated with their nearest multi-antenna AP, while the remaining APs act as interferers. In particular, we have set $N = 4$ antennas per AP serving a single user, i.e., $K = 1$. Also, the imperfect CSI model in that scenario is replaced by the current one while no hardware impairments and channel aging have been assumed. Especially, we have denoted $\bar{\rho}_{\text{tr}}^{\text{sc}} = \bar{\rho}_{\text{tr}}$ and $\bar{p}_{\text{d}}^{\text{sc}} = \frac{N}{K} \bar{p}_{\text{d}}$, where $\bar{\rho}_{\text{tr}}^{\text{sc}}$ and $\bar{p}_{\text{d}}^{\text{sc}}$ are the normalized UL training and DL transmit powers in the case of SCs, in order to guarantee that the total radiated power is equal in both architectures [12]. Clearly, the EE is maximized after a large AP density, being $\lambda_{\text{AP}} = 70$ APs/km², while in the case of CF systems we need only $\lambda_{\text{AP}} = 25$ APs/km².

In Fig. 4, we examine the impact of the SINR constraint γ_0 on the EE. Moreover, we shed light on the tightness of the lower bound on the average SE given by Theorem 1 and an upper bound provided by averaging the instantaneous SE presented by Lemma 1 in terms of MC simulations as in [25]. In particular, we assume that $\gamma_0 \in \{1, 3, 7\}$ to result in an average SE $\log_2(1 + \gamma_0)$ equal to 1, 2, and 3, respectively. Obviously, the EE decreases with γ_0 , which notifies the importance of a target SINR keeping the quality of service in terms of the achievable SE at a satisfactory level according to the specified requirements. Otherwise, we will result in a highly energy-efficient system but useless from the user perspective due to low SE. Furthermore, the gap between the lower and upper bounds is small, which signifies the tightness of the bound proposed by Theorem 1. Hence, the various approximations,

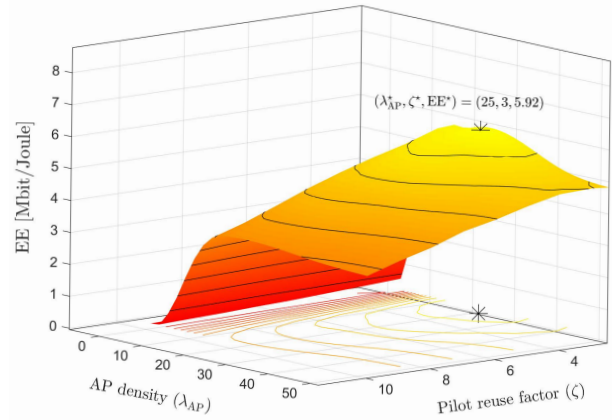


Fig. 2. EE (Mbit/J) of CF mMIMO systems versus the AP density λ_{AP} and pilot reuse factor ζ . The optimal EE is star-marked and the corresponding parameters are provided.

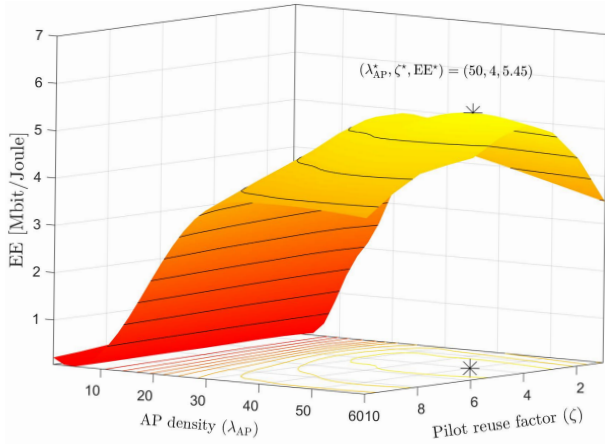
employed for the derivation of this bound, provide reliable results. This observation was expected because the bounding techniques provide tight approximations for a large but finite number of APs, which is the case in CF mMIMO systems. Again, it is depicted that the EE increases with λ_{AP} up to a maximum point, $\lambda_{\text{AP}} = 30$ APs/km² equivalent to the distance among the APs of 103 m approximately, which is reasonable for practical deployments.

In Figs. 5(a) and 5(b), we depict the EE as a function of the ASE for varying γ_0 for $K = 10$ and $K = 20$ users, respectively. From both figures, it is obvious that after a certain value of the ASE the EE decreases, while before, the EE increases together with the ASE. Hence, there are design conditions that could be specified, in order to achieve maximum EE and ASE simultaneously without sacrificing the one over the other. Also, it is shown that these conditions depend on the SINR constraint γ_0 since its increment, being equivalent to an increase of ASE, results in the decrease of the EE as has been already noticed. Regarding the impact of the number of users, we observe that the EE is reduced with K . This is easily shown by (19), which for a given γ_0 , EE is a decreasing function of K because the ASE increases slower than the increase of the APC with the number of users. However, it is shown that we can achieve the optimal EE at higher ASE because increasing the number of users, the ASE increases.

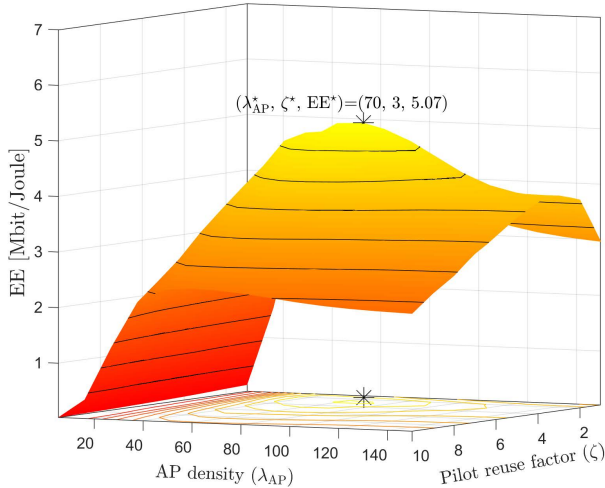
Fig. 6 presents the EE as a function of the number of antennas per AP N and the number of users K when the average SE is equal to 2 Gbit/s, i.e., $\gamma_0 = 3$. Regarding the other parameters under optimization, being the pilot reuse factor and AP density, they are chosen based on Theorems 2 and Theorem 3. Specifically, the optimal values are obtained as $\zeta^* = 3$ and $\lambda_{\text{AP}}^* = 25$. We verify that the EE is a pseudo-concave function of both K and N . Based on simulation, the optimal values are given by $(K^*, N^*) = (5, 16)$ while the corresponding maximum EE is $\text{EE}^* = 6.76$ Mbit/Joule. Notably, these values are confirmed analytically by means Theorems 4 and 5.

TABLE III
PARAMETERS VALUES FOR NUMERICAL RESULTS

| Description | Values | Description | Values |
|-----------------------------|--|-----------------------------|--|
| Number users | $K = 10$ | Boltzmann constant | $\kappa_B = 1.381 \times 10^{-23}$ J/K |
| Number of Antennas/AP | $N = 20$ | Noise temperature | $T_0 = 290$ K |
| AP density | $\lambda_{AP} = 100$ APs/km ² | Noise temperature | $T_0 = 290$ K |
| Communication bandwidth | $W_c = 20$ MHz | Noise figure | $N_F = 9$ dB |
| Carrier frequency | $f_0 = 1.9$ GHz | Fixed power | $P_{FP} = 5$ W |
| Power per pilot symbol | $\rho_{tr} = 100$ mW | Power for AP LO | $P_{LO} = 0.1$ W |
| DL transmit power | $\rho_d = 200$ mW | Power per AP antenna | $P_{AP} = 0.2$ W |
| Path loss exponent | $\alpha = 4$ | Power per UE antenna | $P_{UE} = 0.1$ W |
| Coherence bandwidth | $B_c = 200$ KHz | Power for data coding | $P_{COD} = 0.01$ W/(Gbit/s) |
| Coherence time | $T_c = 1$ ms | Power for data decoding | $P_{DEC} = 0.08$ W/(Gbit/s) |
| UL training duration | $\tau_{tr} = 10$ samples | AP computational efficiency | $L_{AP} = 750$ Gflops/W |
| UL training duration is SCs | $\tau_{tr} = 10$ samples | Power for backhaul traffic | $P_{BT} = 0.025$ W/(Gbit/s) |
| DL training duration is SCs | $\tau_d = 10$ samples | Power amplifier efficiency | $\alpha_{eff} = 0.5$ |



(a) “Cellular” mMIMO systems



(b) SCs systems

Fig. 3. EE (Mbit/J) of versus the AP density λ_{AP} and pilot reuse factor ζ in the cases of a) “cellular” mMIMO systems and b) SCs systems, respectively. The optimal EE is star-marked and the corresponding parameters are provided.

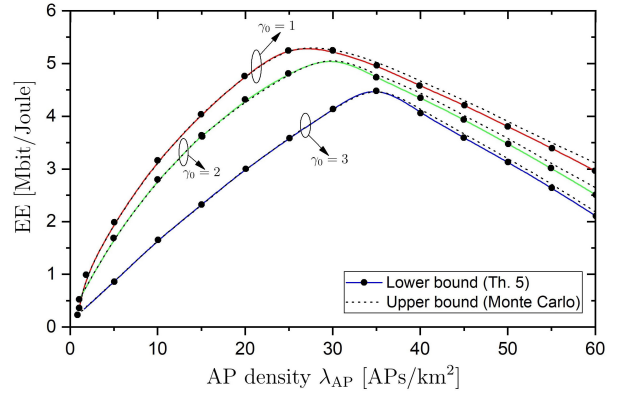
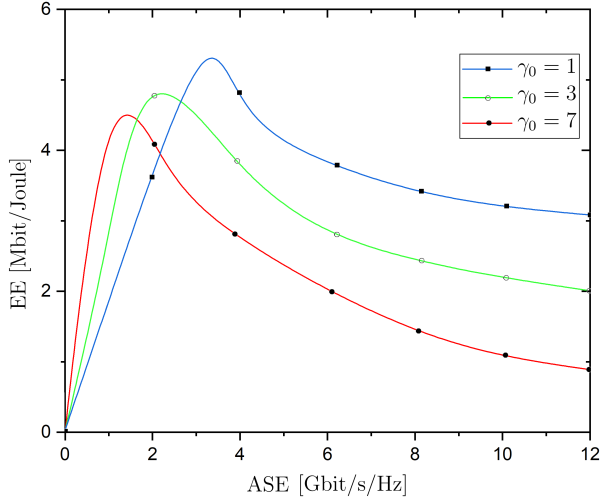


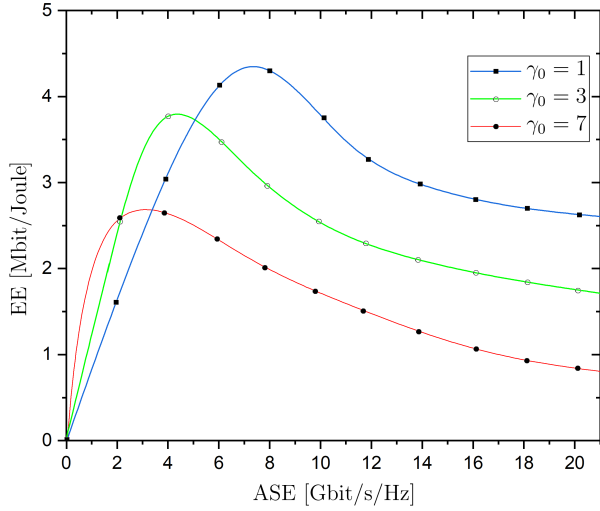
Fig. 4. EE (Mbit/J) of CF mMIMO systems versus the AP density λ_{AP} for different SINR constraints. “Solid-bullet” and “dashed” lines correspond to the lower bound due to the Theorem 1 and upper bound due to MC simulation of the average SE.

VIII. CONCLUSION

Given that network densification is a promising way for high EE, we considered its investigation in a CF mMIMO architecture by assuming both many APs and many antennas per AP. In order to rely on a realistic scenario, we assumed that the APs are PPP distributed. In parallel, we introduced a realistic power consumption model for this setting. Notably, we achieved to derive a new lower bound on the DL average SE for CF mMIMO systems and we provided a novel definition for the ASE which is necessary in the case of CoMP-JT architectures. In this direction, we formulated an EE maximization problem for the DL that enabled the analytical determination of tractable closed-form expressions regarding the optimal EE with respect to the pilot reuse factor, the AP density as well as the number of users and antennas per AP. Remarkably, we achieved to obtain valuable insights concerning the optimization variables. Indeed, the densification in terms of the AP number up to a specific value increases the EE. Moreover, the EE is increased up to a certain point by equipping the APs with more antennas due to the higher array and multiplexing gains that manage to mitigate interference and achieve the resultant higher data rate and lower power consumption.



(a) $K = 10$ users



(b) $K = 20$ users

Fig. 5. EE (Mbit/J) of CF mMIMO systems versus the ASE for different SINR constraints a) $K = 10$ users and b) $K = 20$ users, respectively.

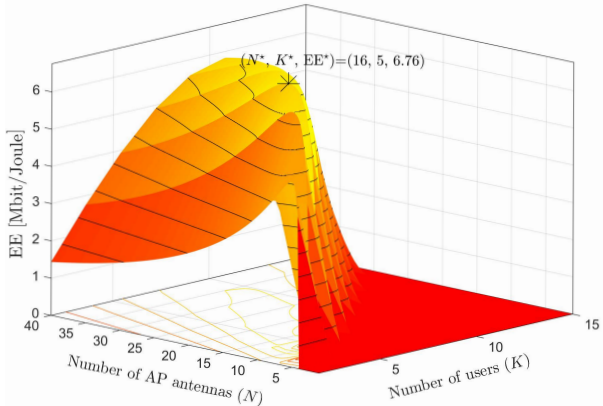


Fig. 6. EE (Mbit/J) of CF mMIMO systems versus the number of AP antennas N and users K . The optimal EE has a black triangle and the corresponding parameters are provided.

APPENDIX A PROOF OF PROPOSITION 1

This proof aims at the derivation of γ_k for finite M by recalling each term of (10). In particular, for the derivation of

each term, we are going to use that $\mathbf{x}^H \mathbf{y} = \text{tr}(\mathbf{y} \mathbf{x}^H)$ for any vectors \mathbf{x} , \mathbf{y} . The term in the numerator becomes

$$\mathbb{E} \left[\mathbf{h}_k^H \mathbf{C}_k \hat{\mathbf{h}}_k \right] = \text{tr} \left(\mathbb{E} \left[\mathbf{C}_k \hat{\mathbf{h}}_k \mathbf{h}_k^H \right] \right) \quad (37)$$

$$= \text{tr} \left(\mathbb{E} \left[\mathbf{L}_k^{-1} \tilde{\mathbf{y}}_k \mathbf{h}_k^H \right] \right) \quad (38)$$

$$= NM, \quad (39)$$

where (38) results after substituting $\hat{\mathbf{h}}_k = \mathbf{L}_k \mathbf{D}^{-1} \tilde{\mathbf{y}}_k$ with $\tilde{\mathbf{y}}_k = [\tilde{\mathbf{y}}_{1k}, \dots, \tilde{\mathbf{y}}_{Mk}]^T$ from (4) while $\mathbf{C}_k = \mathbf{C}^{-1} \mathbf{D} \mathbf{L}_k^{-2}$. The last step is accomplished by applying the expectation between $\tilde{\mathbf{y}}_k$ and \mathbf{h}_k^H . When $i \neq k$, the second-order moment, appearing in the denominator, is written as

$$\mathbb{E} \left[\left| \mathbf{h}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] = \mathbb{E} \left[\left| \hat{\mathbf{h}}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] + \mathbb{E} \left[\left| \tilde{\mathbf{e}}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] \quad (40)$$

$$= \mathbb{E} \left[\left| \hat{\mathbf{h}}_i^H \mathbf{L}_k \mathbf{L}_i^{-1} \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] + \mathbb{E} \left[\left| \tilde{\mathbf{e}}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] \quad (41)$$

$$= N^2 \left(\text{tr}^2(\mathbf{L}_k \mathbf{L}_i^{-1}) + \text{tr}(\mathbf{L}_k^2 \mathbf{L}_i^{-2}) + \text{tr}(\mathbf{C}_i (\mathbf{L}_k - \mathbf{D}^{-1} \mathbf{L}_k^2)) \right) \quad (42)$$

where in (40), we have used that $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{e}}_k$ and the identity $\mathbb{E} [|X + Y|^2] = \mathbb{E} [|X|^2] + \mathbb{E} [|Y|^2]$ holding between two independent random variables with $\mathbb{E} [X] = 0$. In (41), we have applied the property concerning the estimated channels between pilot contaminated users, i.e., $\hat{\mathbf{h}}_k = \mathbf{L}_k \mathbf{L}_i^{-1} \hat{\mathbf{h}}_i$ [20]. The first part of the next equality follows by using [42, Lemma 2], and the second part, not depending on the contamination, results due to the independence between the two random vectors. The last equation is obtained by simple algebraic manipulations. On the contrary, if $i = k$, we have

$$\mathbb{E} \left[\left| \mathbf{h}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] = \mathbb{E} \left[\left| \hat{\mathbf{h}}_k^H \mathbf{C}_k \hat{\mathbf{h}}_k \right|^2 \right] \quad (43)$$

$$= \text{tr}^2 \mathbf{I}_{MN} + \text{tr} \mathbf{I}_{MN} \quad (44)$$

$$= M^2 N^2 + MN, \quad (45)$$

where in (44), we have applied [42, Lemma 2]. In total, we have

$$\mathbb{E} \left[\left| \mathbf{h}_k^H \mathbf{C}_i \hat{\mathbf{h}}_i \right|^2 \right] = N^2 \text{tr}^2 (\mathbf{L}_i^{-1} \mathbf{L}_k) + \begin{cases} N^2 \text{tr} (\mathbf{C}_i \mathbf{L}_k), & i \neq k \\ MN, & i = k. \end{cases} \quad (46)$$

Also, the normalization parameter can be easily written as

$$\begin{aligned} \mu &= \frac{K}{\mathbb{E} \left[\sum_{i=1}^K \hat{\mathbf{h}}_i^H \mathbf{C}_i^2 \hat{\mathbf{h}}_i \right]} \\ &= \left(\frac{N}{K} \sum_{i=1}^K \text{tr} \mathbf{C}_i \right)^{-1} \end{aligned} \quad (47)$$

The proof is concluded by substituting (39), (47) and (46) into (10).

APPENDIX B PROOF OF THEOREM 1

The proof starts with the application of Jensen's inequality that will allow us to derive a tractable lower bound of the DL average SE by moving the expectation inside the logarithm and

continues with the derivation of the expectation of the inverse SINR over the APs instances.

Application of the Jensen inequality to the DL average SE results in

$$\mathbb{E} \left[\log_2 \left(1 + \frac{1}{\tilde{\gamma}_k} \right) \right] \geq \log_2 \left(1 + \frac{1}{\tilde{\gamma}_k} \right), \quad (48)$$

where the expectation applies directly to the inverse SINR $\tilde{\gamma}_k = \mathbb{E} [\tilde{\gamma}_k^{-1}]$.

The inverse SINR provided by (11) can be written as (49) at the top of the next page. where the trace of each matrix is replaced by the sum of its entry-wise elements. For the derivation of the expectation, let a ball of radius R centered at the origin that contains $M = \Phi(B(o, R))$ points with $S = |B(o, R)|$. The first step includes conditioning on this area of radius R and on the number of points in this area. Next, we apply the law of large numbers. Afterwards, we remove the conditioning regarding the number of points, and we assume that the area is infinite, i.e., $R \rightarrow \infty$. Specifically, we have

$$\begin{aligned} \mathbb{E}[\tilde{\gamma}_k^{-1}] &= \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2 N} \sum_{i=1}^K \sum_{m \in \Phi_{\text{AP}} \cap B(o, R)}^M d_m l_{mi}^{-2} \left(N l_{mk} + \frac{1}{K p_d} \right) \right] \\ &+ \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2} \sum_{i \neq k}^K \left(\sum_{m \in \Phi_{\text{AP}} \cap B(o, R)}^M l_{mk} l_{mi}^{-1} \right)^2 \right] \\ &+ \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2} \sum_{m \in \Phi_{\text{AP}} \cap B(o, R)}^M d_m l_{mk}^{-1} \right] + \mathbb{E} \left[\frac{1}{MN} \right], \quad (50) \end{aligned}$$

where in (50), we have let the ball of radius R going to infinity. We continue with the computation of the first term of (50). We have

$$\begin{aligned} \mathcal{I}_1 &= \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2 N} V \right] \\ &= \lim_{R \rightarrow \infty} \mathbb{E}_M \left[\mathbb{E}_{|M} \left[\frac{1}{M^2 N} V \mid M = \Phi(B(o, R)) \right] \right] \quad (51) \end{aligned}$$

$$= \sum_{i=1}^K \lim_{R \rightarrow \infty} \mathbb{E}_M \left[\mathbb{E} \left[M \frac{1}{MN} d_m l_{mi}^{-2} \left(N l_{mk} + \frac{1}{K p_d} \right) \right] \right] \quad (52)$$

$$= \sum_{i=1}^K \mathbb{E} \left[\frac{1}{N} d_m l_{mi}^{-2} \left(N l_{mk} + \frac{1}{K p_d} \right) \right] \quad (53)$$

$$= \mathcal{I}_{11} + \mathcal{I}_{12}, \quad (54)$$

where $V = \sum_{i=1}^K \sum_{m \in \Phi_{\text{AP}} \cap B(o, R)}^M d_m l_{mi}^{-2} \left(N l_{mk} + \frac{1}{K p_d} \right)$, $\mathcal{I}_{11} =$

$\sum_{i=1}^K \mathbb{E} [d_m l_{mi}^{-2} l_{mk}]$, and $\mathcal{I}_{12} = \frac{1}{KN p_d} \sum_{i=1}^K \mathbb{E} [d_m l_{mi}^{-2}]$. In (51), we compute the conditional expectation given the number of points inside the ball, while in (52), one M in the denominator cancels out with the number of points inside the ball, and then, we apply the law of large numbers given the number of APs. Next,

we derive \mathcal{I}_{11} as

$$\mathcal{I}_{11} = \mathbb{E} \left[\sum_{i=1}^K \left(\sum_{j=1}^K |\psi_j \psi_k^H|^2 l_{mj} + \frac{1}{\tau_{\text{tr}} \rho_{\text{tr}}} \right) l_{mi}^{-2} l_{mk} \right] \quad (55)$$

$$= \mathbb{E} \left[\sum_{i=1}^K \sum_{j=1}^K |\psi_j \psi_k^H|^2 l_{mj} l_{mi}^{-2} l_{mk} \right] + \frac{1}{\tau_{\text{tr}} \rho_{\text{tr}}} \mathbb{E} \left[\sum_{i=1}^K l_{mi}^{-2} l_{mk} \right], \quad (56)$$

where the first part of (56) can be written as

$$\begin{aligned} &\mathbb{E} \left[\sum_{i=1}^K \sum_{j=1}^K |\psi_j \psi_k^H|^2 l_{mj} l_{mi}^{-2} l_{mk} \right] \\ &= \begin{cases} \sum_{i=1}^K |\psi_i \psi_k^H|^2 \mathbb{E} [l_{mi}^{-1} l_{mk}] & \text{if } j = i \\ \sum_{i=1}^K \sum_{k=1}^K \mathbb{E} [l_{mi}^{-2} l_{mk}^2] & \text{if } j = k \\ \sum_{j \neq i, k} |\psi_j \psi_k^H|^2 \mathbb{E} [l_{mj} l_{mi}^{-2} l_{mk}] & \text{otherwise} \end{cases} \quad (57) \end{aligned}$$

If $i \neq k$, the expectation in the first branch results in

$$\mathbb{E} [l_{mi}^{-1} l_{mk}] = \mathbb{E} \left[\frac{1}{l_{mi}} \right] \mathbb{E} [l_{mk}] \quad (58)$$

$$\geq \frac{1}{\mathbb{E} [l_{mi}]} \mathbb{E} [l_{mk}] \quad (59)$$

$$= 1, \quad (60)$$

where (58) is obtained due to the independence between the random variables l_{mi} and l_{mk} while (59) has accounted for Jensen's inequality. Notably, (60) is obtained since the two variables have the same marginal distribution. In the condition that $i = k$, the result is the same. Following the same procedure, the expectation in the second branch gives the same result. The expectation in the last branch becomes

$$\mathbb{E} [l_{mj} l_{mi}^{-2} l_{mk}] = \begin{cases} \mathbb{E} [l_{mj} l_{mk}^{-1}] & \text{if } i = k \\ \mathbb{E} [l_{mj} l_{mi}^{-2} l_{mk}] & \text{if } i \neq k \end{cases} \quad (61)$$

The first branch is identical to (58), and gives the same expression. The second branch gives

$$\mathbb{E} [l_{mj} l_{mi}^{-2} l_{mk}] = \mathbb{E} [l_{mj}] \mathbb{E} [l_{mi}^{-2}] \mathbb{E} [l_{mk}] \quad (62)$$

$$\geq \mathbb{E} [l_{mj}] \mathbb{E} [l_{mi}^{-1}]^2 \mathbb{E} [l_{mk}] \quad (63)$$

$$\geq 1, \quad (64)$$

In (62), we have taken into consideration the independence among the variables, and then, in (63) we have applied the inequality $\mathbb{E} [x^2] \geq \mathbb{E} [x]^2$. Eq. (64) is obtained after following similar steps with (60). The second part of (56) is written as

$$\mathbb{E} \left[\sum_{i=1}^K l_{mi}^{-2} l_{mk} \right] = \begin{cases} \mathbb{E} [l_{mi}^{-1}] & \text{if } i = k \\ \sum_{i \neq k} \mathbb{E} [l_{mi}^{-2} l_{mk}] & \text{if } i \neq k \end{cases} \quad (65)$$

Now, the first branch for a general power q becomes

$$\mathbb{E} [l_{mi}^{-q}] \geq \frac{1}{\mathbb{E} [l_{mi}^q]}, \quad (66)$$

where we have applied Jensen's inequality. Note that

$$\mathbb{E} [l_{mi}^q] = 2\pi \left(\int_0^1 y dy + \int_1^\infty y^{-qa+1} dy \right) \quad (67)$$

$$= \frac{q\alpha\pi}{q\alpha - 2}. \quad (68)$$

$$\mathbb{E} [\bar{\gamma}_k^{-1}] = \mathbb{E} \left[\frac{\sum_{i=1}^K \sum_{m=1}^M \left(d_m l_{mi}^{-2} \left(N l_{mk} + \frac{1}{K p_d} \right) \right) + N \sum_{i \neq k} \left(\sum_{m=1}^M l_{mk} l_{mi}^{-1} \right)^2 - N \sum_{m=1}^M d_m l_{mk}^{-1} + M}{M^2 N} \right], \quad (49)$$

Regarding the second branch in (65), we have

$$\mathbb{E} [l_{mi}^{-2} l_{mk}] = \mathbb{E} [l_{mi}^{-2}] \mathbb{E} [l_{mk}] \quad (69)$$

$$\geq \mathbb{E} [l_{mi}^{-1}]^2 \mathbb{E} [l_{mk}] \quad (70)$$

$$\geq \frac{\mathbb{E} [l_{mk}]}{\mathbb{E} [l_{mi}]^2} \quad (71)$$

$$= \frac{1}{\mathbb{E} [l_{mi}]} \quad (72)$$

$$= \frac{\alpha - 2}{\alpha \pi}, \quad (73)$$

where we have applied a property of variance in (70), and the Jensen's inequality in (71). Next, in (73), we have used (68). With respect to the second part of (54) and by following a similar procedure, we have

$$\mathcal{I}_{12} = \frac{1}{KNp_d} \mathbb{E} \left[\sum_{i=1}^K \left(\sum_{j=1}^K |\psi_j \psi_k^H|^2 l_{mj} + \frac{1}{\tau_{tr} \rho_{tr}} \right) l_{mi}^{-2} \right] \quad (74)$$

$$= \frac{1}{\alpha \pi N p_d} \left(\sum_{j=1}^K |\psi_j \psi_k^H|^2 (\alpha - 2) + \frac{\alpha - 1}{\tau_{tr} \rho_{tr}} \right). \quad (75)$$

Substituting the results concerning \mathcal{I}_{11} and \mathcal{I}_{12} , we obtain \mathcal{I}_1 . The second term in (50) becomes

$$\begin{aligned} \mathcal{I}_2 &= \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2} \sum_{i \neq k} U_i \right] \\ &= \lim_{R \rightarrow \infty} \mathbb{E}_M \left[\mathbb{E}_{|M} \left[\frac{1}{M^2} \sum_{i \neq k} U_i \mid M = \Phi(B(o, R)) \right] \right] \quad (76) \end{aligned}$$

$$= \lim_{R \rightarrow \infty} \mathbb{E}_M \left[\mathbb{E}_{|M} \left[\sum_{i \neq k} (\mathbb{E} [l_{mk} l_{mi}^{-1}])^2 \mid M = \Phi(B(o, R)) \right] \right] \quad (77)$$

$$= \lim_{R \rightarrow \infty} \frac{1}{|B(o, R)|} \mathbb{E}_M [M] \sum_{i \neq k} (\mathbb{E} [l_{mk} l_{mi}^{-1}])^2 \quad (78)$$

$$= \lambda_{AP} \sum_{i \neq k} (\mathbb{E} [l_{mk} l_{mi}^{-1}])^2, \quad (79)$$

$$= \lambda_{AP} (K - 1), \quad (80)$$

where $U_i = \left(\sum_{m \in \Phi_{AP} \cap B(o, R)} l_{mk} l_{mi}^{-1} \right)^2$. In (76), we have applied

the law of large numbers, and in (79) we have taken into account that $\mathbb{E}_M [M] = \lambda_{AP} |B(o, R)|$. In (80), we have used similar steps to (60). Similarly, the third term in (50) is obtained

as

$$\mathcal{I}_3 = \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{M^2} \sum_{m \in \Phi_{AP} \cap B(o, R)} d_m l_{mk}^{-1} \right] \quad (81)$$

$$= \mathbb{E} \left[\left(\sum_{j=1}^K |\psi_j \psi_k^H|^2 l_{mj} + \frac{1}{\tau_{tr} \rho_{tr}} \right) l_{mk}^{-1} \right] \quad (82)$$

$$= \sum_{j=1}^K |\psi_j \psi_k^H|^2 + \frac{\alpha - 2}{\alpha \pi \tau_{tr} \rho_{tr}}, \quad (83)$$

where in (82), we have followed similar steps to (52). Regarding the last term in (50), we have

$$\mathcal{I}_4 = \lim_{R \rightarrow \infty} \mathbb{E} \left[\frac{1}{MN} \right] \quad (84)$$

$$\geq \lim_{R \rightarrow \infty} \frac{1}{N \mathbb{E} [M]} \quad (85)$$

$$= \lim_{R \rightarrow \infty} \frac{1}{N \lambda_{AP} |B(o, R)|} \quad (86)$$

$$= 0, \quad (87)$$

where in (85) we have applied Jensen's inequality. Next, we have used that $\mathbb{E}_M [M] = \lambda_{AP} |B(o, R)|$, and we have computed the limit $R \rightarrow \infty$. Substituting \mathcal{I}_1 , \mathcal{I}_2 , \mathcal{I}_3 , and \mathcal{I}_4 into (50), we conclude the proof.

APPENDIX C PROOF OF PROPOSITION 2

The proof, split in two parts, starts with the expression of P_{TX} by means of a lemma, and continues with the presentation of P_{CPC} .

Lemma 3: The total average transmit power consumption due to UL pilot and DL data transmissions of an arbitrary AP is

$$P_{TX} = K \frac{K/\zeta \rho_{tr} + \tau_d \rho_d}{\tau_c}, \quad (88)$$

where $\tau_d = \xi (\tau_c - \tau_{tr})$.

Proof: In each coherence block, each user transmits pilot symbols for a fraction of τ_{tr}/τ_c with power ρ_{tr} , while each AP transmits data symbols for a fraction of τ_d/τ_c with power ρ_d . ■

The second part of (17), concerning the P_{CPC} of an arbitrary AP, is given by [27]

$$P_{CPC} = P_{FP} + P_{TC} + P_{C-BC} + P_{CE} + P_{LP}, \quad (89)$$

where these terms correspond to the power consumptions of circuitry parts. Specifically, P_{FP} expresses the power consumed for site-cooling and control signaling and the traffic-independent mixed power consumption of each backhaul,

P_{TC} for the transceiver chain, P_{C-BC} for coding and load-dependent backhauling cost, while P_{CE} and P_{LP} describe the powers consumed for the processes of channel estimation process and linear processing. Actually, each term depends on the system parameters. Especially, we have that $P_{TC} = NP_{AP} + P_{LO} + KP_{UE}$, where P_{AP} , P_{LO} , and P_{UE} are the powers per AP antenna, AP local oscillator (LO), and the power per user antenna. Moreover, we have $P_{C-BC} = B_w ASE(P_{COD} + P_{DEC} + P_{BT})$, where the terms from left to right denote the bandwidth, the powers for data coding and decoding as well as the total power for the backhaul traffic. Regarding the computation of P_{CE} , we have that the MMSE estimation involves $N\tau_d$ and N operations for the calculations of $\psi_k^H \hat{\mathbf{y}}_m^{tr}$ and $\hat{\mathbf{h}}_{mk}$ in (4) and (5), respectively. In total, the MMSE estimation requires $KN(\tau_{tr} + 1)$ operations needing 3 flops per operation with AP computational efficiency α_{eff} . Given that this procedure takes $\frac{B_w}{\tau_c}$ coherence blocks per second and $\tau_{tr} = \frac{K}{\zeta}$, we have

$$P_{CE} = \frac{3}{L_{AP}} \frac{B_w}{\tau_c} KN \left(\frac{K}{\zeta} + 1 \right). \quad (90)$$

The linear processing power P_{LP} is a result of the powers consumed by precoding/transmitting the data and computation of the precoder, i.e., P_{LP_t} and P_{LP_p} , respectively. Hence, we have

$$P_{LP} = P_{LP_t} + P_{LP_p}, \quad (91)$$

where $P_{LP_t} = \frac{3}{L_{AP}} \frac{B_w}{\tau_c} KN \xi(\tau_c - \tau_{tr})$ with $\tau_{tr} = \frac{K}{\zeta}$, and the power consumed by the conjugate beamformer is given by [8], [27] as $P_{LP_p} = \frac{B_w K}{\tau_c L_{AP}}$. Substituting (88) and the power expressions in (89) into (17), we conclude the proof.

REFERENCES

- [1] A. Papazafeiropoulos *et al.*, "Optimal energy efficiency in cell-free massive MIMO systems: A stochastic geometry approach," in *IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–7.
- [2] J. G. Andrews *et al.*, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.
- [3] Y. Chen *et al.*, "Fundamental trade-offs on green wireless networks," *IEEE Commun. Mag.*, vol. 49, no. 6, pp. 30–37, 2011.
- [4] V. M. Nguyen and M. Kountouris, "Performance limits of network densification," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 6, pp. 1294–1308, June 2017.
- [5] S. Buzzi *et al.*, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 697–709, 2016.
- [6] T. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, no. 11, pp. 3590–3600, November 2010.
- [7] A. K. Papazafeiropoulos and T. Ratnarajah, "Deterministic equivalent performance analysis of time-varying massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 10, pp. 5795–5809, 2015.
- [8] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.
- [9] E. Björnson, J. Hoydis, and L. Sanguinetti, "Massive MIMO has unlimited capacity," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 574–590, 2018.
- [10] E. Björnson *et al.*, "Cooperative multicell precoding: Rate region characterization and distributed strategies with instantaneous and statistical CSI," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4298–4310, 2010.
- [11] J. Gong, S. Zhou, and Z. Zhou, "Networked MIMO with fractional joint transmission in energy harvesting systems," *IEEE Trans. Commun.*, vol. 64, no. 8, pp. 3323–3336, 2016.
- [12] H. Q. Ngo *et al.*, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, 2017.
- [13] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, 2018.
- [14] H. Q. Ngo *et al.*, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Net.*, vol. 2, no. 1, pp. 25–39, 2018.
- [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, 2017.
- [16] A. K. Papazafeiropoulos *et al.*, "Scalable cell-free massive MIMO systems with hardware impairments," in *IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, pp. 1–7.
- [17] P. Parida, H. S. Dhillon, and A. F. Molisch, "Downlink performance analysis of cell-free massive MIMO with finite fronthaul capacity," in *IEEE 88th Vehicular Technology Conference (VTC-Fall)*, pp. 1–6.
- [18] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *arXiv preprint arXiv:1908.03119*, 2019.
- [19] M. Alonzo *et al.*, "Energy-efficient power control in cell-free and user-centric massive MIMO at millimeter wave," *IEEE Trans. Green Commun. Net.*, 2019.
- [20] E. Nayebi *et al.*, "Performance of cell-free massive MIMO systems with MMSE and LSFD receivers," in *IEEE 50th Asilomar Conference on Signals, Systems and Computers*, 2016, pp. 203–207.
- [21] G. Interdonato *et al.*, "Ubiquitous cell-free massive MIMO communications," *EURASIP J. Wireless Commun. Net.*, vol. 2019, no. 1, p. 197, 2019.
- [22] A. K. Papazafeiropoulos *et al.*, "Coverage probability of cell-free massive MIMO systems," in *IEEE International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 1–6.
- [23] E. Björnson and L. Sanguinetti, "Rayleigh fading modeling and channel hardening for reconfigurable intelligent surfaces," *IEEE Wireless Commun. Lett.*, 2020.
- [24] J. G. Andrews, F. Baccelli, and R. K. Ganti, "A tractable approach to coverage and rate in cellular networks," *IEEE Trans. Commun.*, vol. 59, no. 11, pp. 3122–3134, 2011.
- [25] E. Björnson, L. Sanguinetti, and M. Kountouris, "Deploying dense networks for maximal energy efficiency: Small cells meet massive MIMO," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 832–847, 2016.
- [26] H. Ren *et al.*, "Energy efficiency optimization for MIMO distributed antenna systems," *IEEE Trans. Veh. Tech.*, vol. 66, no. 3, pp. 2276–2288, 2017.
- [27] A. Pizzo *et al.*, "Network deployment for maximal energy efficiency in uplink with multislope path loss," *IEEE Trans. Green Commun. Net.*, vol. 2, no. 3, pp. 735–750, 2018.
- [28] L. D. Nguyen *et al.*, "Energy efficiency in cell-free massive MIMO with zero-forcing precoding design," *IEEE Commun. Lett.*, vol. 21, no. 8, pp. 1871–1874, Aug 2017.
- [29] P. Marsch and G. Fettweis, "Uplink CoMP under a constrained backhaul and imperfect channel knowledge," *IEEE Trans. Wireless Commun.*, vol. 10, no. 6, pp. 1730–1742, 2011.
- [30] S. N. Chiu *et al.*, *Stochastic geometry and its applications*. John Wiley & Sons, 2013.
- [31] M. Haenggi, R. K. Ganti *et al.*, "Interference in large wireless networks," *Foundations and Trends® in Networking*, vol. 3, no. 2, pp. 127–248, 2009.
- [32] S. Verdú, *Multuser detection*. Cambridge university press, 1998.
- [33] M. Medard, "The effect upon channel capacity in wireless communications of perfect and imperfect knowledge of the channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 3, pp. 933–946, May 2000.
- [34] E. Björnson, E. G. Larsson, and M. Debbah, "Massive MIMO for maximal spectral efficiency: How many users and pilots should be allocated?" *IEEE Trans. Wireless Commun.*, vol. 15, no. 2, pp. 1293–1308, 2016.
- [35] J. Hoydis, S. ten Brink, and M. Debbah, "Massive MIMO in the UL/DL of cellular networks: How many antennas do we need?" *IEEE J. Select. Areas Commun.*, vol. 31, no. 2, pp. 160–171, February 2013.
- [36] R. Couillet and M. Debbah, *Random matrix methods for wireless communications*. Cambridge University Press, 2011.
- [37] E. Björnson *et al.*, "Optimal design of energy-efficient multi-user MIMO systems: Is massive MIMO the answer?" *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3059–3075, 2015.
- [38] H. Holma and A. Toskala, *LTE for UMTS: Evolution to LTE-Advanced*, Wiley, Ed., 2011.

- [39] L. Welch, "Lower bounds on the maximum cross correlation of signals (corresp.)," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.
- [40] Wolfram Research, Inc., "Mathematica, Version 12.0," champaign, IL, 2019.
- [41] A. Papazafeiropoulos and T. Ratnarajah, "Towards a realistic assessment of multiple antenna HCNs: Residual additive transceiver hardware impairments and channel aging," *IEEE Trans. Veh. Tech.*, vol. 66, no. 10, pp. 9061–9073, Oct 2017.
- [42] E. Björnson, M. Matthaiou, and M. Debbah, "Massive MIMO with non-ideal arbitrary arrays: Hardware scaling laws and circuit-aware design," *IEEE Trans. Wireless Commun.*, vol. 14, no. 8, pp. 4353–4368, Aug 2015.