IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received January 10, 2020, accepted January 23, 2020, date of publication January 27, 2020, date of current version February 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2969815

Access Point Switch ON/OFF Strategies for Green Cell-Free Massive MIMO Networking

GUILLEM FEMENIAS[1], (Senior Member, IEEE), NARJES LASSOUED[2], AND FELIP RIERA-PALOU[1], (Senior Member, IEEE)

[1]Mobile Communications Group, University of the Balearic Islands, 07122 Palma, Spain
[2]INNOV'COM Research Laboratory, University of Carthage, Gabes 6031, Tunisia

Corresponding author: Guillem Femenias (guillem.femenias@uib.es)

This work was funded in part by the Agencia Estatal de Investigación and Fondo Europeo de Desarrollo Regional (AEI/FEDER, UE) under project TERESA (subproject TEC2017-90093-C3-3-R), Ministerio de Economía y Competitividad (MINECO), Spain.

ABSTRACT The combination of user-centric network densification and distributed massive multiple-input multiple-output (MIMO) operation has recently brought along a new paradigm in the wireless communications arena, referred to as cell-free massive MIMO networking. In these networks, a large number of distributed access points (APs), coordinated by a central processing unit (CPU), cooperate to coherently serve a large number of mobile stations (MSs) in the same time/frequency resource. Similar to what has been traditionally done with conventional cellular networks, cell-free massive MIMO networks will be dimensioned to provide the required quality of service (QoS) to MSs under heavy traffic load conditions, and thus they might be underutilized during low traffic load periods, leading to an inefficient use of both spectral and energy resources. Aiming at the implementation of green cell-free massive MIMO networks, this paper proposes and analyzes the performance of different AP switch ON/OFF (ASO) strategies designed to dynamically turn ON/OFF some of the APs based on the number and/or location of the active MSs in the network. The proposed framework considers line-of-sight (LOS) and non-line-of-sight (NLOS) links between APs and MSs, the use of different antenna array architectures at the access points (APs), suitably characterized by array-dependent spatial correlation matrices, and specific power consumption models for APs, MSs and fronthaul links between the APs and the CPU. Numerical results show that the use of properly designed ASO strategies in cell-free massive MIMO networks clearly improve the achievable energy efficiency. Moreover, they also reveal the existing trade-offs among the achievable energy efficiency, the available network-state information, and the hardware configuration (i.e., number of APs, number of transmit antennas per AP, and number of MSs).

INDEX TERMS AP ON/OFF switching, green networking, cell-free massive MIMO, zero-forcing precoding.

I. INTRODUCTION

A. MOTIVATION AND PREVIOUS WORK

Information and communication technologies (ICTs), in general, and wireless communication networks, in particular, have fundamentally and positively changed our way of life. The massive use of ICTs, specially due to the popularization of smart phones and tablets, however, has been steadily increasing the levels of energy consumption, thus significantly contributing to the carbon footprint caused by human activity. In a recent publication [1], Andrae and Edler reported that forecasts for 2030 are that, for the worst-case scenario, ICTs could be responsible for as much as 51% of global energy electricity consumption and contribute up to 23% of the global carbon footprint. With increasing awareness of the potential harmful impact the carbon footprint may have on the environment, it is critical that engineers fully explore innovative greener networking solutions that can meet the growing traffic demand while avoiding the most critical worst-case predictions. Our aim in this paper is to contribute to this goal by proposing the implementation of green cell-free massive multiple-input multiple-output (MIMO) networks.

Cell-free massive MIMO networks have been recently introduced in [2]–[4] as a practical incarnation of the

The associate editor coordinating the review of this manuscript and approving it for publication was Zeeshan Kaleem.

network MIMO concept (also known as coordinated multi-point transmission, distributed MIMO or even cloud RAN) [5]–[8]. In these networks, a massive number of single- or multiple-antenna APs, which are distributed across the coverage area and coordinated by a central processing unit (CPU), cooperate to coherently serve a large number of mobile stations (MSs) in the same time/frequency resource. As in cellular massive MIMO networks [9], channel hardening and favorable propagation conditions are exploited to provide uniformly good quality of service (QoS) to the served MSs using simple MIMO linear signal processing schemes. Inspired by [2]–[4], there has been a great deal of research activity pushing forward the boundaries of this novel wireless network paradigm. Buzzi and Andrea [10], propose a user-centric approach in which each MS is only served by several APs that are selected based on strongest propagation gains. Bashar *et al.* [11] and Femenias and Riera-Palou [12], consider the use of capacity-constrained fronthaul links in sub-6 GHz and millimeter-wave frequency bands, respectively. Nayebi *et al.* [13] and Björnson and Sanguinetti [14], analyze the performance of a cell-free massive MIMO network with zero-forcing (ZF) and minimum mean square error (MMSE) processing, respectively. Particularly related to our work in this paper, Zhang *et al.* [15], Alonzo *et al.* [16], and Bashar *et al.* [17], following the way paved by Ngo *et al.* [4], analyze and optimize the energy efficiency of cell-free massive MIMO networks under different scenarios, including the presence of hardware impairments, the use of millimeter-wave frequency bands or the use of capacity-constrained fronthaul links during the uplink (UL) payload transmission phase. Unfortunately, however, except for the Van Chien's paper, all the aforementioned research works consider a static network scenario in which the number of APs is fixed irrespective of the location of MSs and/or the traffic load they generate, thus obviating the use of one of the most popular energy sustainable paradigms to decrease the carbon footprint of cellular networks, the so-called sleep modes or switch ON/OFF algorithms (see, for instance, [18]–[21] and references therein). As the cellular networks are dimensioned to provide the required QoS to subscribers during the highest load conditions, they might be underused during less busy periods, leading to an inefficient use of both spectral and energy resources. Switch ON/OFF strategies have been traditionally used in these networks to dynamically turn ON/OFF some of the base stations (BSs) based on the location and traffic load generated by the served MSs. User association techniques and cell zooming (also known as cell breathing) strategies are also necessary in such cellular scenarios to complement the use of sleep modes (see, for instance, [22]–[25] and references therein).

B. AIM AND CONTRIBUTIONS

As previously mentioned, and motivated by the above considerations, our main aim in this paper is to address the design and performance evaluation of green cell-free massive MIMO networks based on the use of access point switch-ON/OFF (ASO) strategies. The main contributions of our work can be summarized as follows:

- Mathematically tractable expressions for both the spectral and energy efficiencies of the downlink (DL) and UL payload data transmission phases of a cell-free massive MIMO network with any finite number of APs and MSs are derived. In contrast to most previous works on this topic, these expressions consider the possibility that MSs are in line-of-sight (LOS) with respect to some of the serving APs and in non-line-of-sight (NLOS) with respect to the other ones. Furthermore, the channel model contemplates the use of different antenna array architectures at the APs that can be characterized with suitable spatial correlation matrices. The power consumption model also considers the power consumed at the APs, the MSs and the fronthaul links to and from the CPU. The proposed framework is a non-trivial generalization of previous mathematical models for cell-free massive MIMO networks in [2]–[4], [13], which considered exclusively the propagation through NLOS channels, and also in [26], which contemplated the presence of a LOS but limiting the study to single-antenna APs, thus neglecting spatial correlation effects.
- Taking into account that the problem of selecting the optimal set of APs that must be switched off when serving a given set of MSs is NP-hard, we propose a collection of heuristic suboptimal ASO strategies and discuss their implementation issues as well as their expected complexity versus performance trade-offs. Remarkably, two of the proposed ASO schemes can be considered to provide lower and upper bounds on the energy efficiency performance improvement any sensible AP sleep mode may bring along.
- The substantial energy efficiency benefits produced by the use of ASO strategies are fully quantified by simulating an extensive set of cell-free massive MIMO scenarios. The impact of using different ASO strategies or different antenna configurations at the APs, as well as the repercussion of having cell-free massive MIMO networks comprising different numbers of APs and/or MSs are evaluated. Furthermore, a range of challenging open issues is outlined.

C. PAPER ORGANIZATION AND NOTATIONAL REMARKS

The remainder of this paper is organized as follows. In Section II the proposed green cell-free massive MIMO network is introduced, and different subsections are dedicated to describe the channel model, the UL training phase, and the payload transmission phases for both the DL and the UL. The different performance metrics used in this paper, including the spectral efficiency, the power consumption model and the energy efficiency, are fully developed in Section III. Section IV is devoted to report on the proposed ASO strategies in the context of cell-free massive MIMO. Numerical results and discussions are provided in Section V and, finally, concluding remarks are summarized in Section VI.

Notation: Vectors and matrices are denoted by lower-case and upper-case boldface symbols. The q-dimensional identity matrix is represented by $\boldsymbol{I}_q$. The operator $\|\boldsymbol{X}\|_F$ represents the Frobenius norm of matrix $\boldsymbol{X}$, whereas $\boldsymbol{X}^{-1}$, $\boldsymbol{X}^T$, $\boldsymbol{X}^*$ and $\boldsymbol{X}^H$ denote its inverse, transpose, conjugate and conjugate transpose (also known as Hermitian), respectively. The operators diag($\boldsymbol{x}$) is used to denote a diagonal matrix with the entries of vector $\boldsymbol{x}$ on its main diagonal, and blockdiag($\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$) is used to denote a block diagonal matrix comprising matrices $\boldsymbol{X}_1$, ..., $\boldsymbol{X}_n$ on its main block diagonal. The expectation operator is denoted by $\mathbb{E}\{\cdot\}$. Finally, $\mathcal{CN}(\boldsymbol{m}, \boldsymbol{R})$ denotes a circularly symmetric complex Gaussian vector distributions with mean $\boldsymbol{m}$ and covariance $\boldsymbol{R}$, $\mathcal{N}(0, \sigma^2)$ denotes a real valued zero-mean Gaussian random variable with standard deviation σ, and $\mathcal{U}[a, b]$ represents a random variable uniformly distributed in the range $[a, b]$.

II. SYSTEM MODEL

Let us consider a cell-free massive MIMO network consisting of M randomly distributed APs, each equipped with an array of N antennas and connected to a CPU via an infinite-capacity error-free fronthaul link. Owing to the use of the ASO strategy, each of the APs in the network can be either in active mode (ON) or in sleep mode (OFF). APs in active and sleep modes will be indexed by the sets $\mathcal{M}^A = \{m_1^A, \ldots, m_{M_A}^A\}$ and $\mathcal{M}^S = \{m_1^S, \ldots, m_{M_S}^S\}$, respectively, with $\mathcal{M}^A \cap \mathcal{M}^S = \emptyset$ and $\mathcal{M}^A \cup \mathcal{M}^S = \{1, \ldots, M\}$. The CPU coordinates the communication between the active APs and K geographically distributed single-antenna MSs in the same time-frequency resource. DL and UL transmissions between active APs and MSs are organized in a half-duplex time division duplexing (TDD) operation whereby each coherence interval is split into three phases, namely, the UL training phase, the DL payload data transmission phase and the UL payload data transmission phase. In the UL training phase, all MSs transmit training pilots allowing the APs to estimate the propagation channels to every MS in the network. Subsequently, these channel estimates are used to detect the signals transmitted from the MSs in the UL payload data transmission phase and to compute the precoding filters governing the DL payload data transmission. Obviously, the combined duration/bandwidth of the training, DL and UL phases, denoted as τ_p, τ_d and τ_u, respectively, should not exceed the coherence time/bandwidth of the channel, denoted as τ_c, that is, $\tau_p + \tau_d + \tau_u \leq \tau_c$, with all these intervals specified in samples (or channel uses) on a time-frequency grid.

A. CHANNEL MODEL

As recommended by Björnson and Sanguinetti [14], the typical three-slope pathloss propagation model used in most previous research works on cell-free massive MIMO networking (see, for instance [2]–[4], [12], [13], [27]) will be replaced by a simplified version of the third generation partnership project (3GPP) Urban Microcell model described in [28]. In particular, the link between the mth AP and the kth MS will be considered to be either in LOS or NLOS, with the LOS probability being given by

$$p_{\text{LOS}}(d_{mk}) = \min\left(1, \frac{d_0}{d_{mk}} + \left(1 - \frac{d_0}{d_{mk}}\right) e^{-\frac{d_{mk}}{2d_0}}\right), \quad (1)$$

where d_0 is a reference distance and d_{mk} is the distance between AP m and MS k. The propagation losses (measured in dB) characterizing the propagation link between the mth AP and the kth MS will be modelled as

$$L_{mk} = \alpha + 10\beta \log_{10}(d_{mk}) + \chi_{mk}, \quad (2)$$

where $\chi_{mk} \sim \mathcal{N}\left(0, \sigma_\chi^2\right)$ is the shadow fading component, and the values of parameters α, β and σ_χ depend on whether the corresponding link is in LOS or NLOS. The spatial correlation model for the shadow fading experienced by the different propagation links is described in [3, (54)-(55)].

The resulting uplink channel vector $\boldsymbol{g}_{mk} \in \mathbb{C}^{N\times 1}$ from the kth MS to the mth AP (including both large-scale and small-scale fading) can then be generically characterized as a Ricean fading channel consisting of a LOS component on top of a Rayleigh distributed component modelling the scattered multipath. That is,

$$\boldsymbol{g}_{mk} = \sqrt{\frac{K_{mk}}{K_{mk}+1}}\overline{\boldsymbol{h}}_{mk} + \sqrt{\frac{1}{K_{mk}+1}}\boldsymbol{h}_{mk}, \quad (3)$$

with

$$\overline{\boldsymbol{h}}_{mk} = \alpha_{mk} a^{MS}\left(\overline{\theta}_{mk,1}^{MS}, \overline{\phi}_{mk,1}^{MS}\right)\boldsymbol{a}^{AP}\left(\overline{\theta}_{mk,1}^{AP}, \overline{\phi}_{mk,1}^{AP}\right), \quad (4)$$

and

$$\boldsymbol{h}_{mk} = \sum_{c=1}^{C_{mk}}\sum_{p=1}^{P_{mk}} \alpha_{mk,cp} a^{MS}\left(\theta_{mk,cp}^{MS}, \phi_{mk,cp}^{MS}\right) \times \boldsymbol{a}^{AP}\left(\theta_{mk,cp}^{AP}, \phi_{mk,cp}^{AP}\right), \quad (5)$$

where K_{mk} is the Ricean K-factor, with $K_{mk} = 0$ for NLOS propagation links and $10\log_{10}(K_{mk}) \sim \mathcal{N}\left(\mu_K, \sigma_K^2\right)$ for LOS propagation links. The parameter $\alpha_{mk} = 10^{-L_{mk}/20} e^{j\kappa_{mk}}$, with $\kappa_{mk} \sim \mathcal{U}[0, 2\pi]$, is used to denote the large-scale complex channel gain of the LOS component, C_{mk} and P_{mk} are the number of contributing scattering clusters of the NLOS component and the number of propagation paths per cluster, respectively, $\alpha_{mk,cp}$ is the complex small-scale fading gain on the pth path of cluster c, $a^{MS}\left(\theta_{mk,cp}^{MS}, \phi_{mk,cp}^{MS}\right)$ is the MS antenna element response at the azimuth and elevation angles $\theta_{mk,cp}^{MS}$ and $\phi_{mk,cp}^{MS}$, respectively, and $\boldsymbol{a}^{AP}\left(\theta_{mk,cp}^{AP}, \phi_{mk,cp}^{AP}\right)$ is the normalized array response vector of the AP at the azimuth and elevation angles $\theta_{mk,cp}^{AP}$ and $\phi_{mk,cp}^{AP}$, respectively. As suggested by Akdeniz et al. in [29, Section III.E] (see also [28]), $\theta_{mk,cp}^{MS}$ and $\theta_{mk,cp}^{AP}$ can be generated as wrapped Gaussians around the cluster central angles $\overline{\theta}_{mk,c}^{MS}$ and $\overline{\theta}_{mk,c}^{AP}$ with standard deviation given by the root mean square (rms) azimuth angular spreads for the cluster. Furthermore, $\phi_{mk,cp}^{MS}$ and

TABLE 1. Summary of default simulation parameters.

Parameters	Value
Carrier frequency: f_0	2 GHz
Bandwidth: B	20 MHz
Side of the square coverage area: D	1000 m
AP/MS antenna height: h_{AP}/h_{MS}	10/1.65 m
AP/MS noise figure:	7/9 dB
AP maximum transmit power: P_d	200 mW
MS maximum transmit power: $P_u = P_p$	100 mW
Coherence interval length: τ_c	200 samples
Training phase length: τ_p	20 samples
LOS reference distance: d_0	20 m
Pathloss parameters: α, β, σ_χ	
- Case LOS, $10m < d_{mk} \leq 150m$	34, 2.2, 3
- Case LOS, $d_{mk} > 150m$	-5.17, 4, 3
- Case NLOS	30, 3.67, 4
Shadow fading decorrelation distance: d_{dcorr}	9 m
Shadow fading correlation among APs:	0.5
Ricean K-factor distribution: μ_K, σ_K	9 dB, 5 dB
Number of clusters: C_{mk}	12/19 (LOS/NLOS)
Number of paths per cluster: P_{mk}	20/20 (LOS/NLOS)
Azimuth angular spread (AP): $\lambda_{\text{rms}}^{-1}$	3^o/10^o (LOS/NLOS)
Azimuth angular spread (MS): $\lambda_{\text{rms}}^{-1}$	17^o/22^o (LOS/NLOS)
Elevation angular spread: $\lambda_{\text{rms}}^{-1}$	7^o
Cluster power fraction parameters: r_τ, ζ	3, 4
Power amplifier efficiency: $\alpha_m^{\text{AP}}/\alpha_m^{\text{MS}}$	0.39/0.3
Power per AP/MS/FH traffic: $\xi_m^{\text{AP}}/\xi_k^{\text{MS}}/\xi_m^{\text{FH}}$	0.25/0.25/0.25 $\frac{\text{W}}{\text{Gbps}}$
AP fixed power: $P_{m\,d}^{\text{AP,fix}}/P_{k\,u}^{\text{AP,fix}}$	8/6 W
AP fixed power/RF chain: $P_{m\,d}^{\text{AP,chain}}/P_{m\,u}^{\text{AP,chain}}$	0.2/0.15 W
AP fixed power (sleep): $P_{m\,\text{sleep}}^{\text{AP,fix}}$	0.8 W
AP fixed power/RF chain (sleep): $P_{m\,\text{sleep}}^{\text{AP,chain}}$	0.02 W
MS fixed power: $P_{m\,u}^{\text{MS,fix}}/P_{k\,d}^{\text{MS,fix}}$	1/0.75 W
FH fixed power: $P_m^{\text{FH,fix}}$	5 W
FH fixed power (sleep): $P_{m\,\text{sleep}}^{\text{FH,fix}}$	0.5 W

$\phi_{mk,cp}^{AP}$ can be generated as Laplacians around the cluster central angles $\overline{\phi}_{mk,c}^{MS}$ and $\overline{\phi}_{mk,c}^{AP}$ with scale parameters given by the rms elevation angular spreads for the cluster. The azimuth cluster central angles $\overline{\theta}_{mk,c}^{MS}$ and $\overline{\theta}_{mk,c}^{AP}$ are uniformly distributed in the range $[-\pi, \pi]$ and the elevation cluster central angles $\overline{\phi}_{mk,c}^{MS}$ and $\overline{\phi}_{mk,c}^{AP}$ are set to the corresponding LOS elevation angles. The cluster rms angular spreads are exponentially distributed with a mean equal to $1/\lambda_{\text{rms}}$ that depends on whether we are considering the azimuth or elevation directions. The small-scale scattering fading gains are distributed as

$$\alpha_{mk,cp} \sim \mathcal{CN}\left(0, \gamma_{mk,c} 10^{-L_{mk}/10}\right), \tag{6}$$

where the cluster c is assumed to contribute to the scatter fading with a fraction of power given by

$$\gamma_{mk,c} = \frac{N\gamma'_{mk,c}}{P_{mk}\sum_{j=1}^{C_{mk}}\gamma'_{mk,j}}, \tag{7}$$

with

$$\gamma'_{mk,j} = U_{mk,j}^{r_\tau - 1} 10^{Z_{mk,j}/10}, \tag{8}$$

$U_{mk,j} \sim \mathcal{U}[0,1]$, $Z_{mk,j} \sim \mathcal{N}(0, \zeta^2)$, and the constants r_τ and ζ^2 being treated as model parameters (see [29, Table I] or [28, Table 7.3-6]). Using the channel propagation model just described, the spatial covariance matrix of the scattered multipath component $\boldsymbol{h}_{mk}$ can be obtained as

$$\begin{aligned}\boldsymbol{R}_{mk} &= \mathbb{E}\left\{\boldsymbol{h}_{mk}\boldsymbol{h}_{mk}^H\right\} = 10^{-L_{mk}/10} \\ &\times \sum_{c=1}^{C_{mk}} \gamma_{mk,c} \sum_{p=1}^{P_{mk}} \mathbb{E}\left\{\left|a^{MS}\left(\theta_{mk,cp}^{MS}, \phi_{mk,cp}^{MS}\right)\right|^2\right\} \\ &\times \mathbb{E}\left\{\boldsymbol{a}^{AP}\left(\theta_{mk,cp}, \phi_{mk,cp}\right)\left(\boldsymbol{a}^{AP}\left(\theta_{mk,cp}, \phi_{mk,cp}\right)\right)^H\right\}.\end{aligned} \tag{9}$$

B. SMALL-SCALE TRAINING PHASE: CHANNEL ESTIMATION

Communication in any coherence interval of a TDD-based massive MIMO system invariably starts with the MSs sending the pilot sequences to allow the channel to be estimated at the APs. Let τ_p denote the UL training phase duration (measured in samples on a time-frequency grid) per coherence interval. During the UL training phase, all K MSs simultaneously transmit pilot sequences of τ_p samples to the APs and thus, the $N \times \tau_p$ received UL signal matrix at the mth active AP is given by

$$\boldsymbol{Y}_{p_m} = \sqrt{\tau_p P_p}\sum_{k'=1}^{K}\boldsymbol{g}_{mk'}\boldsymbol{\varphi}_{k'}^T + \boldsymbol{N}_{p_m}, \tag{10}$$

where P_p is the transmit power of each pilot symbol, $\boldsymbol{\varphi}_k$ denotes the $\tau_p \times 1$ training sequence assigned to MS k, with $\|\boldsymbol{\varphi}_k\|_F^2 = 1$, and $\boldsymbol{N}_{p_m} \in \mathbb{C}^{N\times\tau_p}$ is a matrix of independent identically distributed (iid) zero-mean circularly symmetric Gaussian random variables with standard deviation σ_u. Ideally, training sequences should be chosen to be mutually orthogonal, however, since in most practical scenarios it holds that $K > \tau_p$, a given training sequence is assigned to more than one MS, thus resulting in the so-called pilot contamination, a widely studied phenomenon in the context of collocated massive MIMO systems [9], [30], [31].

Considering scenarios where MSs move slowly, it is reasonable to assume that the Ricean K-factors K_{mk}, the LOS components $\bar{\boldsymbol{h}}_{mk}$, and the scatter fading correlation matrices $\boldsymbol{R}_{mk}$ change slowly and can be perfectly known at the mth AP, for all k [26]. Under this assumption, we can define

$$\begin{aligned}\check{\boldsymbol{y}}_{p\,mk} &= \left(\boldsymbol{Y}_{p_m} - \mathbb{E}\left\{\boldsymbol{Y}_{p_m}\right\}\right)\boldsymbol{\varphi}_k^* \\ &= \sum_{k'=1}^{K}\sqrt{\frac{\tau_p P_p}{K_{mk'}+1}}\boldsymbol{h}_{mk'}\boldsymbol{\varphi}_{k'}^T\boldsymbol{\varphi}_k^* + \boldsymbol{N}_{p_m}\boldsymbol{\varphi}_k^*\end{aligned} \tag{11}$$

$$\check{\boldsymbol{g}}_{mk} = \boldsymbol{g}_{mk} - \mathbb{E}\left\{\boldsymbol{g}_{mk}\right\} = \sqrt{\frac{1}{K_{mk}+1}}\boldsymbol{h}_{mk}, \tag{12}$$

and then derive the MMSE estimate for the channel between the kth MS and the mth active AP as [26], [32]

$$\begin{aligned}\hat{\boldsymbol{g}}_{mk} &= \sqrt{\frac{K_{mk}}{K_{mk}+1}}\overline{\boldsymbol{h}}_{mk} \\ &\quad +\mathbb{E}\left\{\check{\boldsymbol{y}}_{p\,mk}\check{\boldsymbol{g}}_{mk}^{H}\right\}\left(\mathbb{E}\left\{\check{\boldsymbol{y}}_{p\,mk}\check{\boldsymbol{y}}_{p\,mk}^{H}\right\}\right)^{-1}\check{\boldsymbol{y}}_{p\,mk} \\ &= \sqrt{\frac{K_{mk}}{K_{mk}+1}}\overline{\boldsymbol{h}}_{mk} + \frac{\sqrt{\tau_p P_p}}{K_{mk}+1}\boldsymbol{R}_{mk}\boldsymbol{\Psi}_{mk}^{-1}\check{\boldsymbol{y}}_{p\,mk}, \quad (13)\end{aligned}$$

where

$$\boldsymbol{\Psi}_{mk} = \tau_p P_p \sum_{k'=1}^{K}\frac{1}{K_{mk'}+1}\boldsymbol{R}_{mk'}\left|\boldsymbol{\varphi}_{k'}^{H}\boldsymbol{\varphi}_{k}\right|^2 + \sigma_u^2\boldsymbol{I}_N. \quad (14)$$

The channel estimate $\hat{\boldsymbol{g}}_{mk}$ and the MMSE channel estimation error $\tilde{\boldsymbol{g}}_{mk} = \boldsymbol{g}_{mk} - \hat{\boldsymbol{g}}_{mk}$ are uncorrelated random vectors distributed as

$$\hat{\boldsymbol{g}}_{mk} \sim \mathcal{CN}\left(\sqrt{\frac{K_{mk}}{K_{mk}+1}}\overline{\boldsymbol{h}}_{mk}, \frac{\tau_p P_p \boldsymbol{R}_{mk}\boldsymbol{\Psi}_{mk}^{-1}\boldsymbol{R}_{mk}^{H}}{(K_{mk}+1)^2}\right), \quad (15)$$

and $\tilde{\boldsymbol{g}}_{mk} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{A}_{mk})$, respectively, where

$$\begin{aligned}\boldsymbol{A}_{mk} &= \mathbb{E}\left\{\tilde{\boldsymbol{g}}_{mk}\tilde{\boldsymbol{g}}_{mk}^{H}\right\} \\ &= \frac{\boldsymbol{R}_{mk}}{K_{mk}+1} - \frac{\tau_p P_p \boldsymbol{R}_{mk}\boldsymbol{\Psi}_{mk}^{-1}\boldsymbol{R}_{mk}^{H}}{(K_{mk}+1)^2}. \quad (16)\end{aligned}$$

C. DOWNLINK PAYLOAD DATA TRANSMISSION

Let us define $\boldsymbol{s}_d = [s_{d1}\ldots s_{dK}]^T$ as the $K \times 1$ vector of symbols jointly (cooperatively) transmitted from the active APs to the MSs, such that $E\left\{\boldsymbol{s}_d\boldsymbol{s}_d^H\right\} = \boldsymbol{I}_K$. Let us also define

$$\boldsymbol{x}_m = \boldsymbol{W}_{d\,m}\boldsymbol{\Upsilon}^{1/2}\boldsymbol{s}_d \quad (17)$$

as the $N \times 1$ vector of signals transmitted from the mth active AP, with $\boldsymbol{W}_{d\,m} = [\boldsymbol{w}_{dm1}\ \ldots\ \boldsymbol{w}_{dmK}] \in \mathbb{C}^{N\times K}$ denoting the precoding matrix at the mth active AP, and $\boldsymbol{\Upsilon} = \mathrm{diag}\,(\boldsymbol{\upsilon}) = \mathrm{diag}\left([\upsilon_1\ \ldots\ \upsilon_K]^T\right)$ being a $K \times K$ diagonal matrix containing the power control coefficients in its main diagonal. These power control coefficients must satisfy the power constraints

$$P_m^{\mathrm{tx}}(\boldsymbol{\upsilon}) = \mathbb{E}\left\{\|\boldsymbol{x}_m\|_F^2\right\} = \sum_{k=1}^{K}\upsilon_k\theta_{mk} \le P_d, \quad (18)$$

for all $m \in \mathcal{M}^A$, where P_d is the maximum average transmit power available at the APs, and we have used the definition

$$\theta_{mk} = \mathbb{E}\left\{\|\boldsymbol{w}_{dmk}\|_F^2\right\}. \quad (19)$$

Using this notation, the signal received by MS k can be expressed as

$$y_{dk} = \sum_{m\in\mathcal{M}^A}\boldsymbol{g}_{mk}^{T}\boldsymbol{x}_m + n_{dk}, \quad (20)$$

where $n_{dk} \sim \mathcal{CN}(0, \sigma_d^2)$. The vector $\boldsymbol{y}_d = \left[y_{d1}\ \ldots\ y_{dK}\right]^T$ containing the signals received by the K scheduled MSs in the network can then be expressed as

$$\boldsymbol{y}_d = \sum_{m\in\mathcal{M}^A}\boldsymbol{G}_m^{T}\boldsymbol{x}_m + \boldsymbol{n}_d = \boldsymbol{G}^T\boldsymbol{W}_d\boldsymbol{\Upsilon}^{1/2}\boldsymbol{s}_d + \boldsymbol{n}_d, \quad (21)$$

where $\boldsymbol{G} = [\boldsymbol{G}_{m_1^A}^T\ \ldots\ \boldsymbol{G}_{m_{M_A}^A}^T]^T$, with $\boldsymbol{G}_m = \left[\boldsymbol{g}_{m1}\ \ldots\ \boldsymbol{g}_{mK}\right]$, represents the equivalent MIMO channel matrix between the K MSs and the M_A active APs, and $\boldsymbol{W}_d = [\boldsymbol{W}_{d\,m_1^A}^T\ \ldots\ \boldsymbol{W}_{d\,m_{M_A}^A}^T]^T$ is the joint precoding filter implemented at the CPU. In particular, using the classical ZF multiuser-MIMO (MU-MIMO) baseband precoder to harness the spatial multiplexing, we have that

$$\boldsymbol{W}_d = \hat{\boldsymbol{G}}^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1} \quad (22)$$

or, equivalently,

$$\boldsymbol{W}_{d\,m} = \hat{\boldsymbol{G}}_m^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1} \quad \forall m \in \mathcal{M}^A, \quad (23)$$

where we have assumed that $\boldsymbol{G} = \hat{\boldsymbol{G}} + \tilde{\boldsymbol{G}}$ and $\boldsymbol{G}_m = \hat{\boldsymbol{G}}_m + \tilde{\boldsymbol{G}}_m$. Consequently, the signal received by the kth MS can be expressed as

$$\begin{aligned}y_{dk} &= \boldsymbol{g}_k^T\hat{\boldsymbol{G}}^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1}\boldsymbol{\Upsilon}^{1/2}\boldsymbol{s}_d + n_{dk} \\ &= \left(\hat{\boldsymbol{g}}_k^T + \tilde{\boldsymbol{g}}_k^T\right)\hat{\boldsymbol{G}}^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1}\boldsymbol{\Upsilon}^{1/2}\boldsymbol{s}_d + n_{dk} \\ &= \sqrt{\upsilon_k}s_{dk} + \tilde{\boldsymbol{g}}_k^T\hat{\boldsymbol{G}}^*\left(\hat{\boldsymbol{G}}^T\hat{\boldsymbol{G}}^*\right)^{-1}\boldsymbol{\Upsilon}^{1/2}\boldsymbol{s}_d + n_{dk}, \quad (24)\end{aligned}$$

where we have defined $\boldsymbol{g_k} = \left[\boldsymbol{g}_{m_1^A k}^T\ \ldots\ \boldsymbol{g}_{m_{M_A}^A k}^T\right]^T$. The first term denotes the useful received signal, the second term contains the interference components due to the use of imperfect channel state information (CSI) (due to UL pilot contamination and noise), and the third term is the thermal noise sample.

D. UPLINK PAYLOAD DATA TRANSMISSION

In the UL, the vector of received signals at the output of the N radio frequency (RF) chains of the mth active AP is given by

$$\begin{aligned}\boldsymbol{r}_{um} &= \sqrt{P_u}\sum_{k'=1}^{K}\boldsymbol{g}_{mk'}\sqrt{\omega_{k'}}s_{uk'} + \boldsymbol{n}_{um} \\ &= \sqrt{P_u}\boldsymbol{G}_m\boldsymbol{\Omega}^{1/2}\boldsymbol{s}_u + \boldsymbol{n}_{um}, \quad (25)\end{aligned}$$

where P_u is the maximum average UL transmit power available at any of the active MSs, $\boldsymbol{s}_u = [s_{u1}\ \ldots\ s_{uK}]^T$ denotes the vector of symbols transmitted by the K active MSs, with $E\left\{\boldsymbol{s}_u\boldsymbol{s}_u^H\right\} = \boldsymbol{I}_K$, $\boldsymbol{\Omega} = \mathrm{diag}\,(\boldsymbol{\omega}) = \mathrm{diag}([\omega_1\ \ldots\ \omega_K]^T)$, with $0 \le \omega_k \le 1$, is a matrix containing the power control coefficients used at the MSs, and $\boldsymbol{n}_{um} \sim \mathcal{CN}(\boldsymbol{0}, \sigma_u^2\boldsymbol{I}_N)$ is the vector of additive thermal noise samples. The received vector

of signals at each of the active APs in the network is forwarded to the CPU via the UL fronthaul links, where they are jointly processed using a set of baseband combining vectors. Assuming the use of ZF MIMO detection, the CPU uses the detection matrix

$$\boldsymbol{W}_u = \left(\hat{\boldsymbol{G}}^H \hat{\boldsymbol{G}}\right)^{-1} \hat{\boldsymbol{G}}^H = \boldsymbol{W}_d^T \tag{26}$$

or, equivalently

$$\boldsymbol{W}_{um} = \left(\hat{\boldsymbol{G}}^H \hat{\boldsymbol{G}}\right)^{-1} \hat{\boldsymbol{G}}_m^H = \boldsymbol{W}_{dm}^T, \quad \forall m \in \mathcal{M}^A, \tag{27}$$

to jointly process the vector $\boldsymbol{r}_u = \left[\boldsymbol{r}_{u\,m_1^A}^T \ \ldots \ \boldsymbol{r}_{u\,m_{M^A}^A}^T\right]^T$ and obtain the vector of detected samples

$$\begin{aligned}\boldsymbol{y}_u &= \boldsymbol{W}_u \boldsymbol{r}_u = \sqrt{P_u}\boldsymbol{W}_u \boldsymbol{G}\boldsymbol{\Omega}^{1/2}\boldsymbol{s}_u + \boldsymbol{\eta}_u \\ &= \sqrt{P_u}\boldsymbol{\Omega}^{1/2}\boldsymbol{s}_u + \sqrt{P_u}\boldsymbol{W}_u \tilde{\boldsymbol{G}}\boldsymbol{\Omega}^{1/2}\boldsymbol{s}_u + \boldsymbol{\eta}_u,\end{aligned} \tag{28}$$

where

$$\boldsymbol{\eta}_u = \boldsymbol{W}_u \boldsymbol{n}_u = \boldsymbol{W}_u \left[\boldsymbol{n}_{u\,m_1^A}^T \ \ldots \ \boldsymbol{n}_{u\,m_{M^A}^A}^T\right]^T. \tag{29}$$

Again, the first term denotes the useful received signal, the second term contains the interference terms due to the use of imperfect CSI, and the third term includes the thermal noise samples. The detected sample corresponding to the symbol transmitted by the kth MS can then be obtained as

$$y_{uk} = \sqrt{P_u}\omega_k^{1/2} s_{uk} + \sqrt{P_u}\left[\boldsymbol{W}_u \tilde{\boldsymbol{G}}\boldsymbol{\Omega}^{1/2}\boldsymbol{s}_u\right]_k + \eta_{uk}, \tag{30}$$

where $[\boldsymbol{x}]_k$ is used to denote the kth entry of vector $\boldsymbol{x}$.

III. MODELING PERFORMANCE METRICS

A. SPECTRAL EFFICIENCY

Analysis techniques similar to those applied, for instance, in [3], [9], [13], [33]–[35], are used in this section to derive DL and UL spectral efficiencies (also known as achievable rates). In particular, the sum of the second and third terms on the right hand side (RHS) of (24), for the DL case, and (30), for the UL case, are treated as effective noise. The additive terms constituting the effective noise are, in both DL and UL cases, mutually uncorrelated, and uncorrelated with s_{dk} and s_{uk}, respectively. Therefore, both the desired signal and the so-called effective noise are uncorrelated. Now, recalling the fact that uncorrelated Gaussian noise represents the worst case, from a capacity point of view, and that the complex-valued fast fading random variables characterizing the propagation channels between different pairs of AP-MS connections are independent, the DL and UL spectral efficiencies (measured in bits per second per Hertz) can be obtained as follows. The DL sum spectral efficiency is given by

$$S_{ed}(\boldsymbol{\upsilon}) = \sum_{k=1}^{K} S_{edk}(\boldsymbol{\upsilon}) = \frac{\tau_d}{\tau_c}\sum_{k=1}^{K}\log_2\left(1+\mathrm{SINR}_{dk}\right), \tag{31}$$

with

$$\mathrm{SINR}_{dk} = \frac{\upsilon_k}{\sum_{k'=1}^{K}\upsilon_{k'}\varpi_{kk'} + \sigma_d^2}, \tag{32}$$

where, using (16), we have that

$$\begin{aligned}\varpi_{kk'} &= \left[\mathrm{diag}\left(\mathbb{E}\left\{\boldsymbol{W}_d^H \mathbb{E}\left\{\tilde{\boldsymbol{g}}_k^* \tilde{\boldsymbol{g}}_k^T\right\}\boldsymbol{W}_d\right\}\right)\right]_{k'} \\ &= \left[\mathrm{diag}\left(\mathbb{E}\left\{\boldsymbol{W}_d^H \boldsymbol{A}_k^* \boldsymbol{W}_d\right\}\right)\right]_{k'},\end{aligned} \tag{33}$$

with $\boldsymbol{A}_k = \mathrm{blockdiag}\left[\boldsymbol{A}_{1k} \ \ldots \ \boldsymbol{A}_{Mk}\right]$. Analogously, the UL sum spectral efficiency is given by

$$S_{eu}(\boldsymbol{\omega}) = \sum_{k=1}^{K} S_{euk}(\boldsymbol{\omega}) = \frac{\tau_u}{\tau_c}\sum_{k=1}^{K}\log_2\left(1+\mathrm{SINR}_{uk}\right), \tag{34}$$

with

$$\mathrm{SINR}_{uk} = \frac{P_u\omega_k}{P_u\sum_{k'=1}^{K}\omega_{k'}\delta_{kk'} + \sigma_{\eta_{uk}}^2}, \tag{35}$$

where

$$\delta_{kk'} = \left[\mathrm{diag}\left(\mathbb{E}\left\{\tilde{\boldsymbol{G}}^H \boldsymbol{w}_{uk}^H \boldsymbol{w}_{uk}\tilde{\boldsymbol{G}}\right\}\right)\right]_{k'} \tag{36}$$

with $\boldsymbol{w}_{uk}$ denoting the kth row of $\boldsymbol{W}_u$, or, equivalently,

$$\begin{aligned}\delta_{kk'} &= \left[\mathrm{diag}\left(\mathbb{E}\left\{\boldsymbol{W}_u \mathbb{E}\left\{\tilde{\boldsymbol{g}}_{k'}\tilde{\boldsymbol{g}}_{k'}^H\right\}\boldsymbol{W}_u^H\right\}\right)\right]_k \\ &= \left[\mathrm{diag}\left(\mathbb{E}\left\{\boldsymbol{W}_u \boldsymbol{A}_{k'} \boldsymbol{W}_u^H\right\}\right)\right]_k,\end{aligned} \tag{37}$$

and

$$\sigma_{\eta_{uk}}^2 = \sigma_u^2\left[\mathrm{diag}\left(\mathbb{E}\left\{\boldsymbol{W}_u\boldsymbol{W}_u^H\right\}\right)\right]_k. \tag{38}$$

B. POWER CONSUMPTION MODEL

As the framework developed in this paper is based on the AP ON/OFF switching strategy, each AP in the network must be either in active or sleep mode. Furthermore, when in active mode, a given AP can be either transmitting during the DL payload data transmission phase or receiving during the UL training and payload data transmission phases. As expected, the power consumed by the mth AP when in the active mode depends on the radiated power P_m^{tx} during the DL payload data transmission phase, or on the UL spectral efficiency $S_{eu}(\boldsymbol{\omega})$ during the UL payload data transmission phase. However, it also depends on parameters such as the efficiency of the power amplifier, the small-signal RF transceiver power, the baseband power, the feeder loses, the DC-DC power supply loses, the main supply losses, or the cooling losses [36]–[40]. In the sleep mode, the AP is in a reduced power consumption state in which it is not completely turned off and can then be readily activated. Although the AP is not radiating or receiving power when in the sleep mode, there are components such as the power supply, some of the signal processing blocks, and part of the cooling system that are still active and thus consuming power. Consequently, the total power consumption of the mth AP can be approximated by

a linear model as follows (see, for instance, [36]–[40] and references therein)

$$P_m^{\mathrm{AP}} = \begin{cases} \dfrac{P_m^{\mathrm{tx}}(\boldsymbol{\upsilon})}{\alpha_m^{\mathrm{AP}}} + P_{m\,d}^{\mathrm{AP,fix}} + NP_{m\,d}^{\mathrm{AP,chain}} & \text{DL Active} \\ B\xi_m^{\mathrm{AP}} S_{eu}(\boldsymbol{\omega}) + P_{m\,u}^{\mathrm{AP,fix}} + NP_{m\,u}^{\mathrm{AP,chain}} & \text{UL Active} \\ P_{m\,\mathrm{sleep}}^{\mathrm{AP,fix}} + NP_{m\,\mathrm{sleep}}^{\mathrm{AP,chain}} & \text{Sleep,} \end{cases} \tag{39}$$

where α_m^{AP} is the power amplifier efficiency at the mth AP, B is the system bandwidth, ξ_m^{AP} is the traffic-dependent power consumption coefficient (in Watt per bit/s), $P_{m\,d}^{\mathrm{AP,fix}}$ and $P_{m\,u}^{\mathrm{AP,fix}}$ denote, respectively, the DL and UL power consumption figures that are independent of both the number of RF chains and the traffic load, $P_{m\,d}^{\mathrm{AP,chain}}$ and $P_{m\,u}^{\mathrm{AP,chain}}$ model the DL and UL traffic-independent power consumed by the circuitry related to each RF chain of the mth AP, respectively and, finally, $P_{m\,\mathrm{sleep}}^{\mathrm{AP,fix}}$ and $P_{m\,\mathrm{sleep}}^{\mathrm{AP,chain}}$ are the RF chain-independent and RF chain-dependent power consumed by the mth AP when in sleep mode.

A similar power consumption model can be established for the fronthaul links connecting the APs to the CPU. In particular, the power consumed by the mth fronthaul link when in active mode depends on the amount of traffic it has to convey and, thus, the total power consumption can be approximated as [4], [27]

$$P_m^{\mathrm{FH}} = \begin{cases} B\xi_m^{\mathrm{FH}} S_{ed}(\boldsymbol{\upsilon}) + P_m^{\mathrm{FH,fix}} & \text{DL Active} \\ B\xi_m^{\mathrm{FH}} S_{eu}(\boldsymbol{\omega}) + P_m^{\mathrm{FH,fix}} & \text{UL Active} \\ P_{m\,\mathrm{sleep}}^{\mathrm{FH},fix} & \text{Sleep,} \end{cases} \tag{40}$$

where ξ_m^{FH} is the traffic-dependent power consumption coefficient (in Watt per bit/s), $P_m^{\mathrm{FH,fix}}$ is the traffic-independent power consumption when in active mode, and $P_{m\,\mathrm{sleep}}^{\mathrm{FH}}$ accounts for the power consumed by the mth fronthaul link when in sleep mode.

The power consumption model for the MSs can also be approximated as

$$P_k^{\mathrm{MS}} = \begin{cases} B\xi_k^{\mathrm{MS}} S_{edk}(\boldsymbol{\upsilon}) + P_{k\,d}^{\mathrm{MS,fix}} & \text{DL} \\ \dfrac{P_u \omega_k}{\alpha_k^{\mathrm{MS}}} + P_{k\,u}^{\mathrm{MS,fix}} & \text{UL,} \end{cases} \tag{41}$$

where, again, α_k^{MS} is the power amplifier efficiency at the kth MS, ξ_k^{MS} is the traffic-dependent power consumption coefficient (in Watt per bit/s), $P_{m\,d}^{\mathrm{MS,fix}}$ and $P_{m\,u}^{\mathrm{MS,fix}}$ model the power consumed by the internal circuitry of the MS independently of the average radiated power, and $S_{edk}(\boldsymbol{\omega})$ denotes the DL spectral efficiency of the kth MS.

Putting all the pieces together, the total power consumption of the cell-free massive-MIMO network can be modeled as

$$P_{T\,d}(\boldsymbol{\upsilon}) = P_{Td}^{\mathrm{fix}} + B\sum_{k=1}^{K} \xi_k^{\mathrm{MS}} S_{edk}(\boldsymbol{\upsilon}) + \sum_{m\in\mathcal{M}^A} \left(\frac{\tau_d}{\tau_c} \frac{P_m^{\mathrm{tx}}(\boldsymbol{\upsilon})}{\alpha_m^{\mathrm{AP}}} + B\xi_m^{\mathrm{FH}} S_{ed}(\boldsymbol{\upsilon}) \right), \tag{42}$$

for the DL payload data transmission phase, and as

$$P_{T\,u}(\boldsymbol{\omega}) = P_{Tu}^{\mathrm{fix}} + \sum_{k=1}^{K} \frac{\tau_u}{\tau_c} \frac{P_u \omega_k}{\alpha_m^{\mathrm{MS}}} + B \sum_{m\in\mathcal{M}^A} \left(\xi_m^{\mathrm{AP}} + \xi_m^{\mathrm{FH}} \right) S_{eu}(\boldsymbol{\omega}), \tag{43}$$

for the UL payload data transmission phase, with

$$P_{Tl}^{\mathrm{fix}} = \frac{\tau_l}{\tau_c} \left[\sum_{k=1}^{K} P_{k\,l}^{\mathrm{MS,fix}} + \sum_{m\in\mathcal{M}^A} \left(P_m^{\mathrm{FH,fix}} + P_{m\,l}^{\mathrm{AP,fix}} + NP_{m\,l}^{\mathrm{AP,chain}} \right) + \sum_{m\in\mathcal{M}^S} \left(P_{m\,\mathrm{sleep}}^{\mathrm{FH},fix} + P_{m\,\mathrm{sleep}}^{\mathrm{AP,fix}} + NP_{m\,\mathrm{sleep}}^{\mathrm{AP,chain}} \right) \right] \tag{44}$$

where l has been used as a token to represent either the DL ($l = d$) or the UL ($l = u$). As stated by Desset *et al.* [38], although this simple linear model is not designed to provide very accurate absolute figures, it will enable a fair comparison among different ON/OFF switching strategies for green cell-free massive-MIMO networking.

C. ENERGY EFFICIENCY

The energy efficiency during the DL and UL payload data transmission phases can be expressed as

$$E_{ed}(\boldsymbol{\upsilon}) = \frac{BS_{ed}(\boldsymbol{\upsilon})}{P_{T\,d}(\boldsymbol{\upsilon})} \tag{45}$$

and

$$E_{eu}(\boldsymbol{\omega}) = \frac{BS_{eu}(\boldsymbol{\omega})}{P_{T\,u}(\boldsymbol{\omega})}, \tag{46}$$

respectively. We can also define a weighted energy efficiency metric as

$$E_e(\boldsymbol{\upsilon}, \boldsymbol{\omega}) = \mu E_{ed}(\boldsymbol{\upsilon}) + (1-\mu) E_{eu}(\boldsymbol{\omega}), \tag{47}$$

where $0 \leq \mu \leq 1$ is a weighting coefficient allowing for the control of a trade-off between DL and UL energy efficiencies.

IV. AP SWITCH-ON/OFF STRATEGIES

In the context of green cell-free massive MIMO networking, the ultimate objective of optimal ASO strategies is to select M_A out of M APs in such a way that the resulting energy efficiency for a given arrangement of MSs is maximized. On the one hand, finding such an optimal subset of APs is an NP-hard problem, thus requiring of the evaluation of the performance provided by all possible combinations of M_A out of M APs. Hence, assuming that the number of APs in a cell-free massive MIMO network, by its very nature, is large, this selection will call for the development of heuristic suboptimal algorithms. On the other hand, under ideal conditions, the set of selected APs should be adapted to scenario variations due, among others, to changes

in the location of the MSs and/or the geographical distribution of shadow fading. In particular, Chien *et al.* [41] have very recently proposed an energy-efficient cell-free massive MIMO scheme aiming at minimizing the total DL power consumption at the APs assuming that some of them can be turned off. Critically, the proposed strategy implies solving a computationally-intensive non-convex optimization problem at the rate of change of the large-scale parameters of the system. In most practical scenarios, however, these variations occur too quickly so as to allow the implementation of such high-complexity adaptive selection schemes. In the following we describe some heuristic ASO strategies and comment on their possible implementation issues as well as on their expected complexity versus performance trade-offs. Starting with a pure random selection ASO scheme (Subsection IV-A), described for lower-bound benchmarking purposes, we then propose two strategies (Subsections IV-C and IV-D) that are loosely based on similar techniques used in conventional wireless networks. Finally, three novel proposals are presented in Subsections IV-B, IV-E and IV-F that have been specially tailored for the cell-free massive MIMO network deployment considered in this paper.

A. RANDOM SELECTION ASO

Simple random AP switch-ON/OFF (or random switching) is probably the most straightforward AP selection scheme. Using this strategy, each of the available APs is equally likely to be put in sleep mode and, therefore, the only parameter to be adjusted when maximizing the energy efficiency of the network will be the number of APs that must be put in sleep mode as a function of the number (or spatial density) of MSs that have to be served. Being completely unaware of the possible effects switching off a particular AP may have on the global performance of the network, the random selection ASO (RS-ASO) strategy is expected to provide a lower bound on the energy efficiency performance improvement any sensible ASO may bring along.

B. MIXTURE DISCREPANCY-BASED GREEDY ASO

Assuming that the MSs are uniformly distributed on the coverage area or, equivalently, under a complete ignorance about the spatial distribution of MSs, keeping the locations of the set of active APs as uniform as possible seems to be advantageous in terms of spectral/energy efficiency of a cell-free massive MIMO network. This is basically because a uniform spatial distribution of APs tends to match the uniform statistical distribution of MSs.

In the field of statistics, discrepancy has gained much popularity as a tool to measure the deviation between the empirical and the theoretical uniform distribution (see, for instance, [42] and references therein). Examples of discrepancies that have been suggested as possible measures of uniformity are, among many others, the star L_p-discrepancy ($p \neq 2$), the star L_∞-discrepancy, the generalized, centered and symmetrical L_2-discrepancies, the discrete discrepancy, the Lee discrepancy or the mixture discrepancy. As concluded by Zhou *et al.* [43], among all of these suggested measures of uniformity, the mixture discrepancy they introduce is the one fulfilling the major quantity of desirable mathematical and computational properties to construct uniform designs.

For a given set $\mathcal{M}^A$ of active APs located at positions $\boldsymbol{P}(\mathcal{M}^A) = \left\{\boldsymbol{p}_{m_1^A}, \ldots, \boldsymbol{p}_{m_{M_A}^A}\right\}$, with $\boldsymbol{p}_m = (p_{m1}, p_{m2}) \in \mathbb{R}^2$, the corresponding analytical expression of mixture discrepancy can be obtained by using [43, eq. (18), (with $n = M_A$, $s = 2$, and $\boldsymbol{x}_m = \boldsymbol{p}_m$)]. Hence, using this particular criterion, among all the sets of active APs with a given cardinality, the optimal one to serve a set of K MSs uniformly distributed over the service area would be the one showing the minimum mixture discrepancy. Furthermore, the optimal set cardinality (i.e., the optimal number of active APs) would be the one providing the maximum energy efficiency. Having a large number of APs in the network, that is, having a large M, NP-hardness forbids the implementation of a brute force algorithm to solve this optimization problem. Consequently, the iterative mixture discrepancy-based greedy ASO (MD-ASO) algorithm is proposed that, starting with a set containing all the APs in the network, in each iteration switches-off the single AP producing the highest decrease in the mixture discrepancy metric. Note that, again, the optimal number of active APs (under the greedy strategy) when serving a given amount of MSs would be the one providing the maximum energy efficiency.

C. SPATIAL REGULARITY-BASED GREEDY ASO

Aiming at providing a high energy efficiency while maintaining good user satisfaction, ASO strategies can also increase the number of APs in sleep mode while keeping the locations of the active APs as regular as possible. Defining perfect regularity as the case in which APs are placed on a triangular lattice [44], metrics are needed to quantify the spatial regularity of the different sets of (potential) active APs. A particularly interesting metric that was already used by Lagum *et al.* [45] to design cell switch-off algorithms is the geometry-based metric of spatial regularity [46], [47]. Given a set $\mathcal{M}^A$ of active APs located at positions $\boldsymbol{P}(\mathcal{M}^A)$, it is defined as

$$C_D\left(\mathcal{M}^A\right) = \frac{\sigma_D\left(\mathcal{M}^A\right)}{k_D\,\mu_D\left(\mathcal{M}^A\right)}, \tag{48}$$

where $\mu_D\left(\mathcal{M}^A\right)$ and $\sigma_D\left(\mathcal{M}^A\right)$ are the mean and the standard deviation of the Delaunay edge lengths of the triangulation between the points representing the positions in $\boldsymbol{P}(\mathcal{M}^A)$, respectively, and $k_D = 0.492$ is a normalization factor ensuring that, on average, the geometry-based metric of spatial regularity for a Poisson point process is equal to 1 [47]. Note, also, that $C_D\left(\mathcal{M}^A\right) = 0$ when the active APs are located on a perfectly regular triangular lattice. Again, as NP-hardness forbids the use of brute force algorithms to find the subset $\mathcal{M}^A$ showing the highest spatial regularity (i.e., the minimum geometry-based metric of spatial regularity $C_D(\mathcal{M}^A)$), and with a cardinality maximizing the energy efficiency of

the network, the iterative spatial regularity-based greedy ASO (SR-ASO) algorithm is proposed that, starting with a set containing all the APs in the network, in each iteration switches-off the single AP whose transition to sleep mode produces the highest increase in the spatial regularity metric of the remaining active APs. The optimal number of active APs (under the greedy strategy) when serving a given amount of MSs would be, once more, the one providing the maximum energy efficiency.

D. NEAREST NEIGHBOUR-BASED ASO

Again, under the assumption of a random uniform distribution of MSs on the coverage area, the distribution shown by the set of active APs will tend to match that of the MSs if, at each step, out of the two APs that are nearer to each other, the one that is nearer to a third one is put in sleep mode. This strategy, that will be termed as the nearest neighbour-based ASO (NN-ASO) scheme and was previously used by Lagum et al. in [45], generates solutions maximizing the minimum pairwise distance among the set of active APs. The optimal number of active APs when serving a given amount of MSs would be the one maximizing the energy efficiency.

E. PROPAGATION LOSSES-AWARE ASO

Assuming a uniform spatial distribution of MSs over the service area, already described ASO strategies are only able to adapt the number of active APs and/or the components of the set of active APs to large-scale variations of traffic demand (i.e., long-term variations of K). If, instead, the proposed cell-free massive MIMO network is allowed to use time-dynamic ASO strategies able to adapt to shorter-term traffic variations, a new door is open to improve the energy efficiency while maintaining good service coverage and user satisfaction. In particular, even though the long-term spatial distribution of MSs can be assumed to be uniform, the dynamics of users and shadowing at a shorter time scale give rise to short-term spatial distributions of users that could benefit from a distribution of active APs adapted to them. The propagation losses-aware ASO (PL-ASO) strategy, based on the availability of large-scale propagation losses between APs and MSs, aims at such an adaptive behaviour by switching-off those APs showing large propagation losses to the served MSs. The set $\mathcal{M}^A$ under the PL-ASO strategy is comprised of M_A active APs that are selected using two different procedures depending on the particular values of M_A and K. On the one hand, for those cases in which $M_A \geq K$, the first selected active APs are those exhibiting the minimum propagation losses to each of the K MSs. Note that in this first step, the number of APs added to $\mathcal{M}^A$ is less or equal than K. In a second step, after removing the already selected APs from the set of selectable ones, the procedure is repeated. That is, the algorithm selects (in an ordered manner) the APs, out of the remaining ones, whose propagation losses to each of the K MSs is minimum. This procedure is repeated until the number of selected APs is equal to M_A. On the other hand, for those cases in which $M_A < K$, use is made of the k-means clustering method to partition the K large-scale propagation vectors $\boldsymbol{\beta}_k = [\beta_{1k} \ \ldots \ \beta_{Mk}]$ into M_A clusters in which each propagation vector belongs to the cluster with the nearest centroid, serving as a representative of the cluster [48]. Now, using the virtual propagation losses vectors characterizing the M_A centroids, the same selection procedure previously described is applied to select the M_A active APs. Again, the optimal number of active APs under this criterion would be the one maximizing the energy efficiency for a given network load.

F. OPTIMAL ENERGY EFFICIENCY-BASED GREEDY ASO

The only way to obtain an upper bound on the performance any practical ASO strategy can bring along is to evaluate, for each possible value of M_A, all the available combinations of active APs (i.e., $\binom{M}{M_A}$ sets of active APs) and select the one providing the maximum energy efficiency. As was previously mentioned, however, this is an NP-hard problem of a huge computational complexity. In order to obtain an approximation to this upper bound, a greedy algorithm can be applied that, in the first iteration, starts with the M available APs, evaluates the M possible configurations of $(M-1)$ active APs resulting from switching off one of them, and selects the configuration maximizing the energy efficiency. In the second iteration, the same procedure is repeated but, in this case, starting with the $(M-1)$ APs selected in the first step in order to greedily choose the best configuration of $(M-2)$ active APs. The same operations are repeated in the following iterations until obtaining the configuration of active APs maximizing the energy efficiency of the network. This will be termed as the optimal energy efficiency-based greedy ASO (OG-ASO) strategy and will be used as a benchmark against which the performance of the other ASO schemes will be assessed.

V. NUMERICAL RESULTS

In this section, numerical results are provided to quantitatively evaluate the performance of the proposed ASO strategies in terms of its energy efficiency. Replicating the scenario typically used in most of the relevant literature on this topic (see, for instance, [2]–[4], [12], [13], [27]), APs and MSs are uniformly distributed at random within a square coverage area of size $D \times D$ m^2. Boundary effects are avoided by wrapping around this square area at the edges, thus simulating the effects of operating a network with an infinite coverage area.

Default parameters used to set-up the simulation scenarios under evaluation in the following subsections are summarized in Table 1 and are inspired by a variety of prior research works (see, for instance, [3], [14], [27], [28], [36], [39] and references therein). Furthermore, although the proposed analytical framework can be applied assuming any of the power control strategies previously proposed in the literature (see, for instance, [3], [4], [13]), results presented in this paper have been obtained using the heuristic solution proposed by Nayebi et al. in [13, eq. (21)]

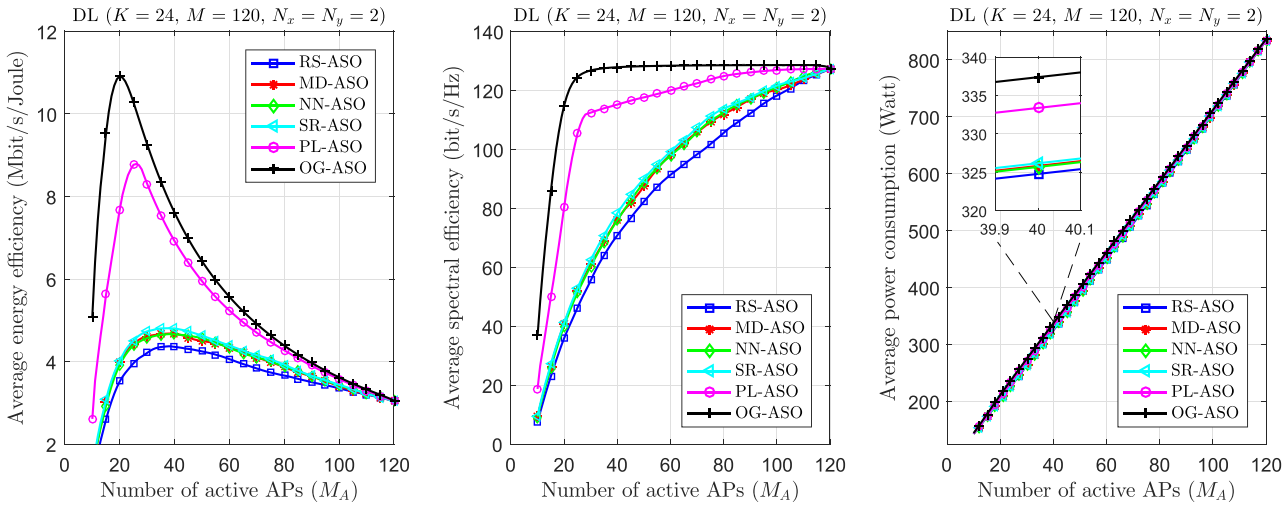

FIGURE 1. Impact of the ASO strategy on the DL average energy efficiency, espectral efficiency and power consumption as a function of the number of active APs.

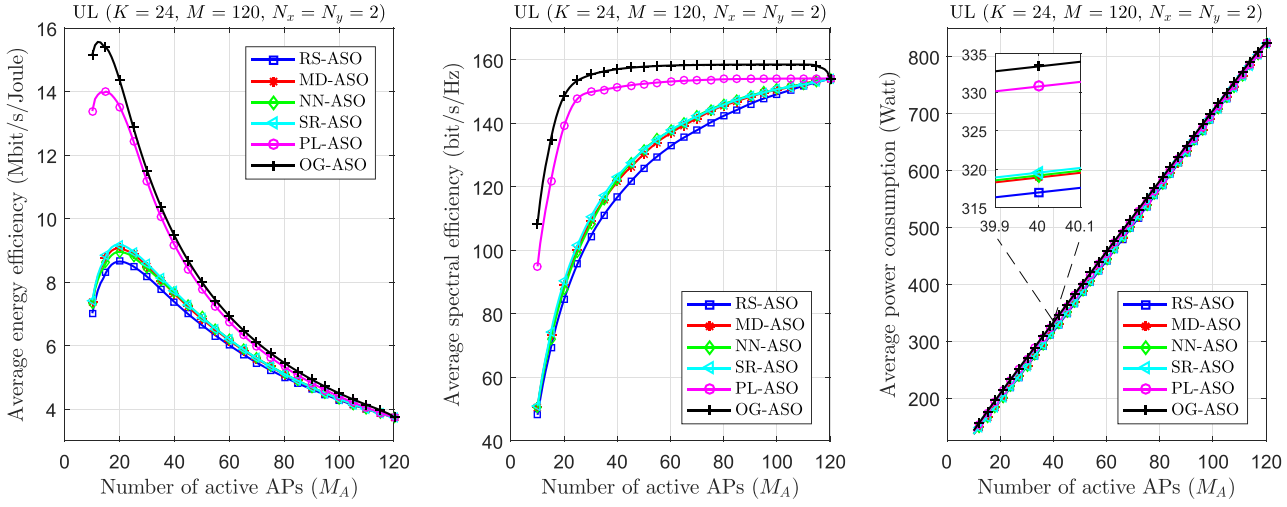

FIGURE 2. Impact of the ASO strategy on the uplink average energy efficiency, espectral efficiency and power consumption as a function of the number of active APs.

(i.e., $\upsilon_k = P_d / \left(\max_m \sum_{k'=1}^{K} \theta_{mk'}\right)$ for all k) for the DL case, and the full-power transmission strategy (i.e., $\omega_k = 1$ for all k) for the UL case. Furthermore, the balanced random pilot assignment scheme has been applied [12], where MSs are allocated pilot sequences that are sequentially and cyclically selected from the ordered set of available orthogonal pilots.

A. IMPACT OF THE ASO STRATEGY

Our aim in this subsection is to assess the performance of the proposed ASO strategies in terms of energy/spectral efficiencies and power consumption. Accordingly, the energy efficiency, spectral efficiency and power consumption versus the number of active APs is presented in Figs. 1 and 2 for each of the propounded schemes and for both the DL (i.e., $\mu = 1$) and the UL (i.e., $\mu = 0$), respectively. All results have been obtained assuming the default system parameters described in Table 1, and the availability of $M = 120$ APs equipped with 2×2 uniform planar arrays (UPAs) of vertical half-wave dipoles located on a half-wave grid and serving $K = 24$ MSs. The first important result to note from these figures is that, irrespective of the number of active APs in the network or the ASO strategy under use, the energy efficiency of the UL is much higher than that provided by the DL. This is basically due to two main reasons. Firstly, although both transmission segments show very similar average power consumption metrics, the fixed power consumption in the UL is considerably lower than that in the DL. Secondly, the use of full power transmission in the UL provides a clear advantage, in terms of spectral efficiency, with respect to the constrained power control transmission implemented in the DL. Using a max-min power control approach would lead to almost identical spectral efficiency performance results for both the

DL and the UL (see, for instance, results presented in [12]) and, in this case, the energy efficiency advantage shown by the UL segment would only be due to the lower fixed power consumption.

Another interesting result disclosed in Figs. 1 and 2 is that the energy-efficiency achieved by the RS-ASO and OG-ASO schemes act, respectively, as lower- and upper-bounds on the performance attained by any of the other proposed ASO strategies. In fact, the proposed ASO schemes can be classified in four groups as a function of the system state information they manage. The RS-ASO scheme would be the only member in the first group, comprising those ASO strategies that are completely unaware of the network state and thus making blind AP switch-off decisions. The ASO strategies in the second group, comprising the MD-ASO, the SR-ASO and the NN-ASO schemes, are all based on the assumption that the MSs are uniformly distributed on the service coverage area and make only use of very large-scale system-state information: the geographical location of the APs. Being only aware of such a poor network-state information, it is not surprising that the energy-efficiency performance improvement they may offer with respect to the pure RS-ASO algorithm is rather meager when compared to that achieved by the idealistic OG-ASO scheme. Still, the performance improvements offered by these strategies are not negligible at all and, furthermore, as it can be observed in Figs. 1 and 2, the achievable energy-efficiency increases as more APs are switched-off until it reaches a maximum that, for this particular number of MSs and irrespective of the ASO strategy under consideration, is located around $M_A = 37$ active APs for the DL case and $M_A = 20$ active APs for the UL case. Switching-off a greater number of APs would produce a worsening of both the energy and spectral efficiencies of the system. Comparing the performance metrics achieved by these ASO strategies, it is quite evident that the MD-ASO and NN-ASO schemes behave similarly and are outperformed by the SR-ASO scheme. The only ASO strategy in the third group is the PL-ASO scheme that, based on the knowledge of the large-scale propagation losses between APs and MSs, dynamically adapts to short-term variations of the spatial distribution of MS and, as shown in Figs. 1 and 2, definitely outperforms the ASO strategies in the first and second groups. Again, the energy efficiency provided by this strategy increases when switching-off some of the APs in the cell-free massive MIMO network, and a maximum is obtained, for this particular scenario, when $M_A = 25$ active APs for the DL case and $M_A = 15$ for the UL case. Finally, the fourth group of proposed ASO strategies, only comprising the OG-ASO scheme, assumes the complete knowledge of all long-term network-state information necessary to calculate the achievable energy-efficiency, including, among others, the channel spatial correlation matrices, the power control matrices or the power consumption metrics. The energy-efficiency performance gap between this rather idealistic approach and the much simpler PL-ASO is not excessively wide but, remarkably, the maximum performance

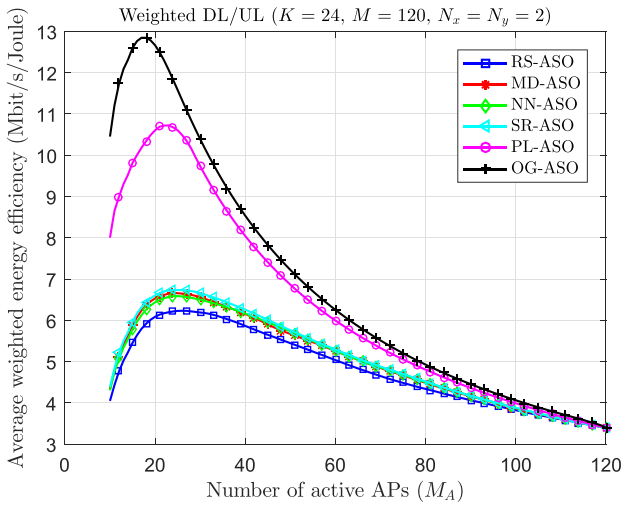

FIGURE 3. Impact of the ASO strategy on the average (equally) weighted energy efficiency as a function of the number of active APs.

level is achieved with $M_A = 19$ active APs for the DL case and $M_A = 12$ for the UL case, that is, with a number of active APs that is considerably less than the number of MSs in the service area (recall that the APs are equipped with 2×2 UPAs).

Fig. 3 shows the impact of the ASO strategy under use on the average weighted energy efficiency of a cell-free massive MIMO network using a DL/UL weighting coefficient $\mu = 0.5$ (i.e., both the DL and the UL are given the same importance). Note that, in this case, the DL/UL weighting operation makes the resulting energy efficiency metric taking intermediate values lying between those obtained when both links where dealt separately. Because of that, the number of active APs optimizing the average weighted energy efficiency are also located somewhere in between those resulting from the pure DL and UL optimizations. Specifically, in this particular scenario, the optimal number of active APs necessary to serve $K = 24$ MSs is M_A =26, 25, 24, 24, 23 and 17 APs for the RS-ASO, NN-ASO, MD-ASO, SR-ASO, PL-ASO and OG-ASO strategies, respectively.

Among all the proposed ASO strategies, the most adequate to be implemented in a cell-free massive MIMO scenario, based on the use of very large-scale network-state information, would be the SR-ASO, as it is the one providing the best performance versus complexity/implementability trade-off. Hence, convinced that the conclusions drawn by using this strategy would be qualitatively equivalent to those that could be drawn by using any of the other ASO schemes, results presented in the next subsections will be obtained assuming the use of SR-ASO. For similar reasons, we will only consider the optimization of the DL segment (i.e., $\mu = 1$).

B. IMPACT OF THE ANTENNA CONFIGURATION AT THE APs

A zoomed view of the DL achievable energy efficiency is plotted in Fig. 4 against the number of active APs in a

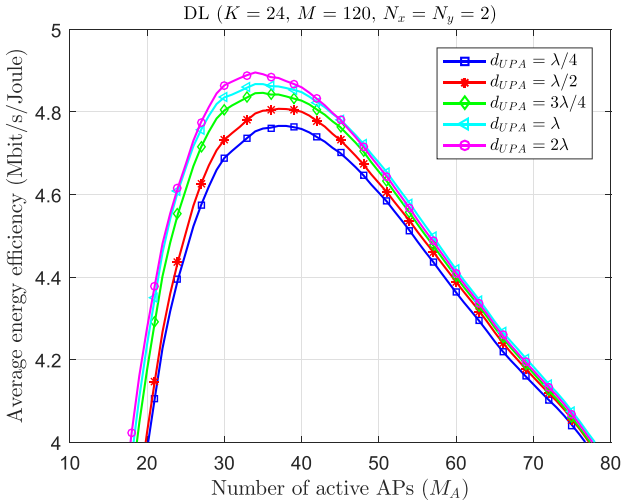

FIGURE 4. Impact of the antenna separation (2 × 2 UPA of vertical half-wave dipoles) on the average energy efficiency as a function of the number of active APs.

cell-free massive MIMO network of $M = 120$ APs equipped with 2×2 UPAs of vertical half-wave dipoles and serving $K = 24$ MSs. In order to evaluate the impact of antenna separation (i.e., spatial correlation), the distance between the neighbour vertical half-wave dipoles takes values in the set $\{\lambda/4, \lambda/2, 3\lambda/4, \lambda, 2\lambda\}$, where λ is the wavelength of the carrier signal. As expected, results show that increasing the distance between antennas reduces spatial correlation and results in non negligible energy efficiency increments. Furthermore, increasing the distance between antennas also reduces the number of active APs needed to maximize the achievable energy efficiency. In particular, the optimal number of active APs goes from $M_A = 37$ for $d_{\text{UPA}} = \lambda/4$ to $M_A = 34$ for $d_{\text{UPA}} = 2\lambda$. Nevertheless, it is worth stressing that, keeping all the other parameters constant, the marginal increment of performance produced by each new increment in antenna separation suffers from the law of diminishing returns and thus, antenna separations on the order of $d_{\text{UPA}} = \lambda$ seem to be quite reasonable from the point of view of the energy efficiency performance versus antenna array size trade-off and, consequently, this is the default antenna separation that will be used from this point onwards.

In order to assess the impact of the array antenna configuration on the performance of the proposed cell-free massive MIMO system, the energy efficiency, spectral efficiency and power consumption versus the number of active APs is presented in Figs. 5 and 6 for different number of transmit antennas, and assuming the use of either a UPA or a uniform linear array (ULA) of vertical half-wave dipoles, respectively. Again, all results have been obtained assuming the default system parameters described in Table 1, the use of a RS-ASO strategy, and the availability of $M = 120$ APs serving $K = 24$ MSs. As it can be observed, irrespective of whether use is made of a UPA or a ULA, both the average spectral efficiency and the power consumption increase with the number of transmit antennas. The growth patterns of both metrics as the number of transmitting antennas increases, however, are very different. On the one hand, the spectral efficiency is clearly subject to the law of diminishing returns, since the larger the number of transmit antennas in the array, the lower the spectral efficiency increase produced by the addition of more antennas to the array. On the other hand, there is a part of the power consumption metric that increases linearly with the number of transmit antennas constituting the array. As a consequence of these dissimilar growth patterns, the average energy efficiency of small arrays can be improved by increasing the number of transmit antennas but, further increasing the number of transmit antennas of an already large array will only slightly increase the average spectral efficiency at the cost of decreasing the average energy efficiency. Interestingly, however, increasing the number of transmit antennas reduces the number of active APs necessary

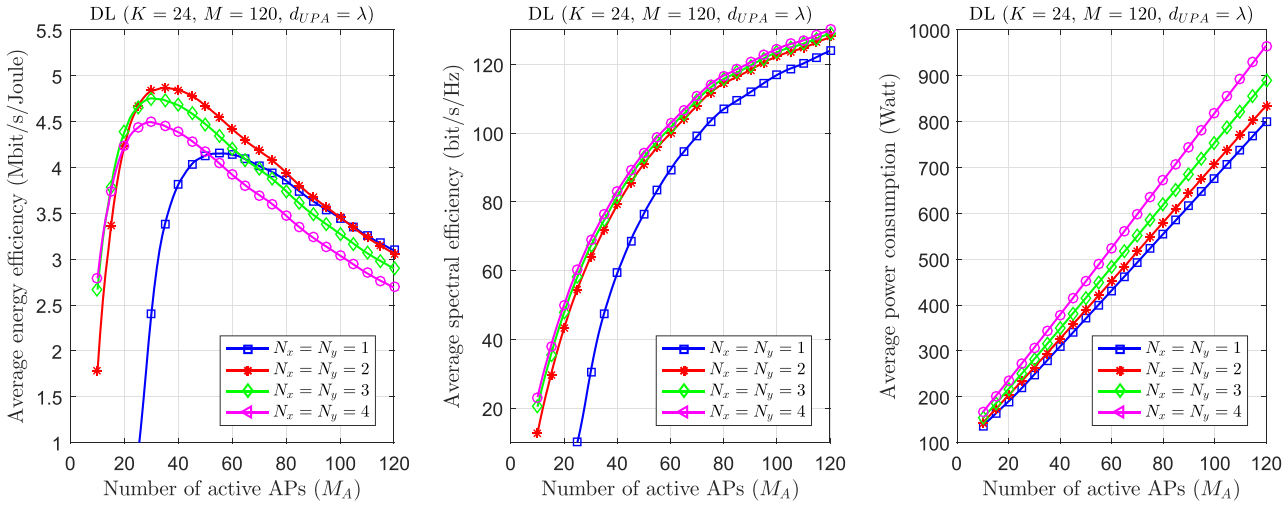

FIGURE 5. Impact of the UPA configuration on the DL average energy efficiency, espectral efficiency and power consumption as a function of the number of active APs.

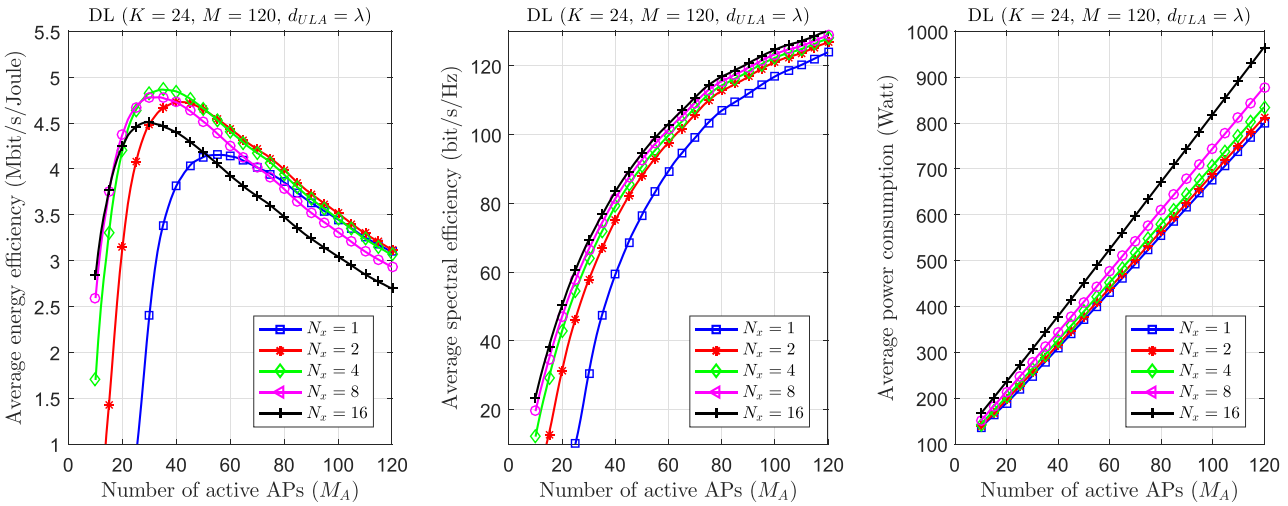

FIGURE 6. Impact of the ULA configuration on the DL average energy efficiency, espectral efficiency and power consumption as a function of the number of active APs.

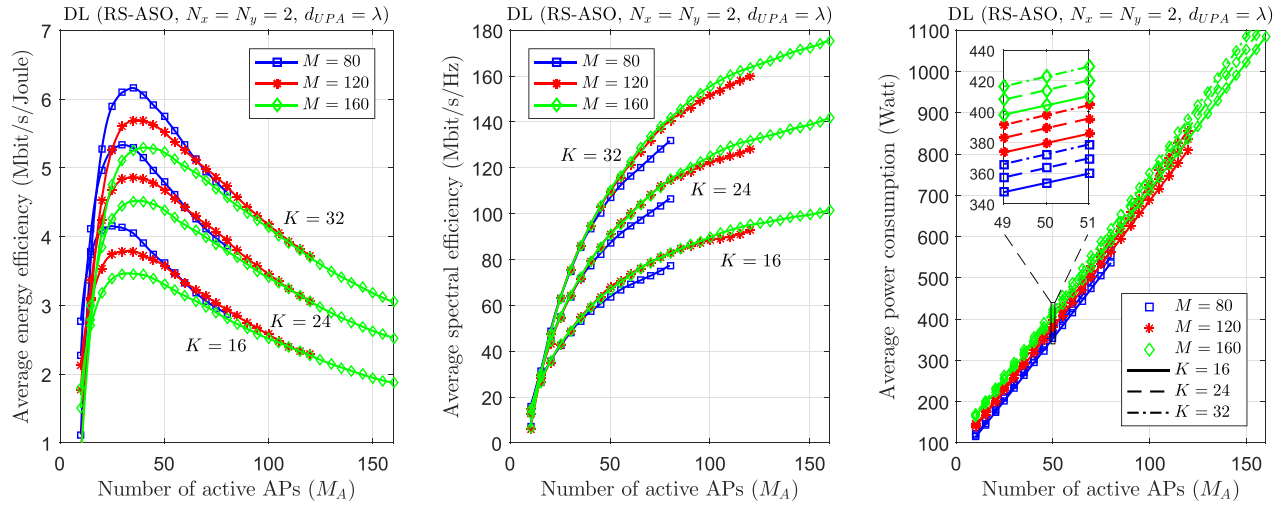

FIGURE 7. Impact of the ULA configuration on the DL average energy efficiency, espectral efficiency and power consumption as a function of the number of active APs.

TABLE 2. Impact of antenna configuration.

Array class	$N_x \times N_y$	M_A^*	$E_{ed}^*(\boldsymbol{\upsilon})$ [Mbps/J]
Single antenna	1×1	56	4.15
UPA	2×2	34	4.87
	3×3	30	4.75
	4×4	29	4.49
ULA	2×1	41	4.73
	4×1	34	4.86
	8×1	30	4.78
	16×1	29	4.51

to maximize the average spectral efficiency for that specific configuration. In particular, for the scenario under consideration (i.e., RS-ASO, $\mu = 1$, $M = 120$ APs and $K = 24$ MSs), the optimal average energy efficiency $E_{ed}^*(\boldsymbol{\upsilon})$ and optimal number of active APs M_A^* are summarized in Table 2.

C. IMPACT OF THE NUMBER OF APs AND MSs IN THE NETWORK

As shown in Fig. 7, increasing either the number of APs or the number of MSs in the network results in an increase in both the average spectral efficiency and the average power consumption. The increments induced on these performance metrics, however, given their relative magnitudes, produce completely different effects on the average energy efficiency of the network. In particular, it can be observed that, for given numbers of active APs and MSs, having more APs in the network (i.e., a larger M value) results in average power consumption increments that are larger than those experienced by the average spectral efficiency and, as a consequence, the average energy efficiency decreases. On the contrary, for fixed numbers of both available and active APs, increasing the number of MSs translates into a small increase

in power consumption and a considerable improvement of spectral efficiency, thus resulting in a substantial amelioration of the average energy efficiency of the network. Therefore, the conclusion seems quite obvious, if high energy efficiency is to be achieved, it is very important that the number of both the available and active APs be appropriately adapted to the number of MSs that have to be serviced.

Another interesting result worth mentioning is that the optimal number of active APs, M_A^*, increases with both the number of available APs and the number of MSs in the network. In particular, in an scenario with $K = 24$ MSs and $M = 80$, 120 or 160 APs, the optimal number of active APs is equal to $M_A^* = 30$, 34 or 36, respectively, whereas in an scenario with $M = 120$ APs and $K = 16$, 24 or 32 MSs, the optimal number of active APs is equal to $M_A^* = 32$, 34 or 37, respectively.

VI. CONCLUSION

A novel analytical framework for the performance analysis of green cell-free massive MIMO networks has been introduced in this paper. The proposed framework considers the use of different ASO strategies designed to dynamically turn ON/OFF some of the APs based on the number of active MSs (i.e., traffic load) in the network. In particular, six ASO strategies have been proposed: the pure random selection scheme, denoted as RS-ASO, three selection strategies aiming at keeping the locations of the set of active APs as uniform as possible, denoted as MD-ASO, NN-ASO and SR-ASO, a selection strategy that exploits the availability of time-dynamic information about short-term traffic variations, denoted as PL-ASO and, finally, a greedy optimal selection strategy, denoted as OG-ASO. Generalizing the system models described in most previous research works on this topic, the proposed analytical framework considers that MSs can be in LOS with respect to some of the serving APs and in NLOS with respect to the other ones. Furthermore, a channel model contemplating the use of different antenna array configurations at the APs has been applied. Additionally, a realistic power consumption model has been proposed for the cell-free massive MIMO architecture that takes into account the power consumed at the APs, the MSs and the fronthaul links to and from the CPU. Numerical results have shown that, irrespective of the number of active MSs in the network or the ASO strategy under use, the energy efficiency of the UL is higher than that provided by the DL. Furthermore, they also reveal that the energy-efficiency achieved by the RS-ASO and OG-ASO schemes act, respectively, as lower- and upper-bounds on the performance attained by any of the other proposed ASO strategies. Among all the proposed ASO strategies, the most adequate to be implemented in a cell-free massive MIMO scenario, based on the use of very large-scale network-state information, would be the SR-ASO, as it is the one providing the best performance versus complexity/implementability trade-off. Moreover, increasing the number of transmit antennas of small antenna arrays can serve to improve the energy efficiency of the network but, further increasing the number of transmit antennas of an already large array will only slightly increase the average spectral efficiency at the cost of decreasing the average energy efficiency. Interestingly, however, for any particular antenna configuration, increasing the number of transmit antennas reduces the number of active APs necessary to maximize the average spectral efficiency. Finally, it is worth mentioning that the energy efficiency of a cell-free massive MIMO network can only be maximized whenever both the number of available APs and the number of active APs are suitably adapted to the number of served MSs.

As for the specific research directions in the future, there are many open issues still remaining, including, among many others, the implementation of more sophisticated ASO strategies, the consideration of non-uniform long-term user distributions, the evaluation of more intricate power control strategies, the study of the impact that the use of millimeter wave frequency bands may produce on the achievable energy efficiency, the consideration of finite-capacity fronthaul links between the APs and the CPU, or the assessment of scalability aspects related to the green nature of the proposed approach.

REFERENCES

[1] A. Andrae and T. Edler, "On global electricity usage of communication technology: Trends to 2030," *Challenges*, vol. 6, no. 1, pp. 117–157, Apr. 2015.

[2] 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. IEEE 16th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Jun. 2015, pp. 201–205.

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

[4] 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.

[5] M. Karakayali, G. Foschini, and R. Valenzuela, "Network coordination for spectrally efficient communications in cellular systems," *IEEE Wireless Commun.*, vol. 13, no. 4, pp. 56–61, Aug. 2006.

[6] D. Gesbert, S. Hanly, H. Huang, S. Shamai Shitz, O. Simeone, and W. Yu, "Multi-cell MIMO cooperative networks: A new look at interference," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1380–1408, Dec. 2010.

[7] R. Irmer, H. Droste, P. Marsch, M. Grieger, G. Fettweis, S. Brueck, H.-P. Mayer, L. Thiele, and V. Jungnickel, "Coordinated multipoint: Concepts, performance, and field trial results," *IEEE Commun. Mag.*, vol. 49, no. 2, pp. 102–111, Feb. 2011.

[8] A. Checko, H. L. Christiansen, Y. Yan, L. Scolari, G. Kardaras, M. S. Berger, and L. Dittmann, "Cloud RAN for mobile networks—A technology overview," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 405–426, 1st Quart., 2015.

[9] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of Massive MIMO*. Cambridge, U.K.: Cambridge Univ. Press, 2016.

[10] 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.

[11] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, and M. Debbah, "Cell-free massive MIMO with limited backhaul," 2018, *arXiv:1801.10190*. [Online]. Available: https://arxiv.org/abs/1801.10190

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

[13] 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.

[14] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," 2019, *arXiv:1903.10611*. [Online]. Available: http://arxiv.org/abs/1903.10611
[15] J. Zhang, Y. Wei, E. Bjornson, Y. Han, and S. Jin, "Performance analysis and power control of cell-free massive MIMO systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55302–55314, 2018.
[16] 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.
[17] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "Energy efficiency of the cell-free massive MIMO uplink with optimal uniform quantization," *IEEE Trans. Green Commun. Netw.*, vol. 3, no. 4, pp. 971–987, Dec. 2019.
[18] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.
[19] F. Han, S. Zhao, L. Zhang, and J. Wu, "Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems," *IEEE Access*, vol. 4, pp. 4959–4973, 2016.
[20] P. Gandotra, R. K. Jha, and S. Jain, "Green communication in next generation cellular networks: A survey," *IEEE Access*, vol. 5, pp. 11727–11758, 2017.
[21] N. Piovesan, A. F. Gambin, M. Miozzo, M. Rossi, and P. Dini, "Energy sustainable paradigms and methods for future mobile networks: A survey," *Comput. Commun.*, vol. 119, pp. 101–117, Apr. 2018.
[22] H. Tabassum, U. Siddique, E. Hossain, and M. J. Hossain, "Downlink performance of cellular systems with base station sleeping, user association, and scheduling," *IEEE Trans. Wireless Commun.*, vol. 13, no. 10, pp. 5752–5767, Oct. 2014.
[23] C. Jia and T. J. Lim, "Resource partitioning and user association with sleep-mode base stations in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 7, pp. 3780–3793, Jul. 2015.
[24] X. Xu, C. Yuan, W. Chen, X. Tao, and Y. Sun, "Adaptive cell zooming and sleeping for green heterogeneous ultradense networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 2, pp. 1612–1621, Feb. 2018.
[25] H. Jiang, S. Yi, L. Wu, H. Leung, Y. Wang, X. Zhou, Y. Chen, and L. Yang, "Data-driven cell zooming for large-scale mobile networks," *IEEE Trans. Netw. Service Manag.*, vol. 15, no. 1, pp. 156–168, Mar. 2018.
[26] H. Q. Ngo, H. Tataria, M. Matthaiou, S. Jin, and E. G. Larsson, "On the performance of cell-free massive MIMO in Ricean fading," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 980–984.
[27] 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.
[28] *Study on 3D Channel Model for LTE (Release 12)*, document TR 36.873 (Version 12.7.0), 3GPP, Dec. 2017.
[29] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1164–1179, Jun. 2014.
[30] 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.
[31] O. Elijah, C. Y. Leow, T. A. Rahman, S. Nunoo, and S. Z. Iliya, "A comprehensive survey of pilot contamination in massive MIMO—5G system," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 905–923, 2nd Quart., 2016.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.
[33] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links?" *IEEE Trans. Inf. Theory*, vol. 49, no. 4, pp. 951–963, Apr. 2003.
[34] H. Yang and T. L. Marzetta, "Capacity performance of multicell large-scale antenna systems," in *Proc. 51st Annu. Allerton Conf. Commun., Control, Comput. (Allerton)*, Oct. 2013, pp. 668–675.
[35] G. Interdonato, H. Q. Ngo, E. G. Larsson, and P. Frenger, "On the performance of cell-free massive MIMO with short-term power constraints," in *Proc. IEEE 21st Int. Workshop Comput. Aided Model. Design Commun. Links Netw. (CAMAD)*, Oct. 2016, pp. 225–230.
[36] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.
[37] S. Tombaz, P. Monti, K. Wang, A. Vastberg, M. Forzati, and J. Zander, "Impact of backhauling power consumption on the deployment of heterogeneous mobile networks," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2011, pp. 1–5.
[38] C. Desset, B. Debaillie, V. Giannini, A. Fehske, G. Auer, H. Holtkamp, W. Wajda, D. Sabella, F. Richter, M. J. Gonzalez, H. Klessig, I. Godor, M. Olsson, M. A. Imran, A. Ambrosy, and O. Blume, "Flexible power modeling of LTE base stations," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2012, pp. 2858–2862.
[39] E. Bjornson, 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, Apr. 2016.
[40] B. Dai and W. Yu, "Energy efficiency of downlink transmission strategies for cloud radio access networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 1037–1050, Apr. 2016.
[41] T. V. Chien, E. Björnson, and E. G. Larsson, "Optimal design of energy-efficient cell-free massive MIMO: Joint power allocation and load balancing," 2019, *arXiv:1911.11375*. [Online]. Available: https://arxiv.org/abs/1911.11375
[42] E. Androulakis, K. Drosou, C. Koukouvinos, and Y.-D. Zhou, "Measures of uniformity in experimental designs: A selective overview," *Commun. Statist.-Theory Methods*, vol. 45, no. 13, pp. 3782–3806, Jul. 2016.
[43] Y.-D. Zhou, K.-T. Fang, and J.-H. Ning, "Mixture discrepancy for quasi-random point sets," *J. Complex.*, vol. 29, nos. 3–4, pp. 283–301, Jun. 2013.
[44] B. Rengarajan, G. Rizzo, and M. Ajmone Marsan, "Energy-optimal base station density in cellular access networks with sleep modes," *Comput. Netw.*, vol. 78, pp. 152–163, Feb. 2015.
[45] Q.-N. Le-The, T. Beitelmal, F. Lagum, S. S. Szyszkowicz, and H. Yanikomeroglu, "Cell switch-off algorithms for spatially irregular base station deployments," *IEEE Wireless Commun. Lett.*, vol. 6, no. 3, pp. 354–357, Jun. 2017.
[46] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous traffic distribution in heterogeneous wireless cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2252–2265, Oct. 2015.
[47] F. Lagum, S. S. Szyszkowicz, and H. Yanikomeroglu, "CoV-based metrics for quantifying the regularity of hard-core point processes for modeling base station locations," *IEEE Wireless Commun. Lett.*, vol. 5, no. 3, pp. 276–279, Jun. 2016.
[48] F. Riera-Palou, G. Femenias, A. G. Armada, and A. Perez-Neira, "Clustered cell-free massive MIMO," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–6.

GUILLEM FEMENIAS (Senior Member, IEEE) received the Telecommunication Engineer degree and the Ph.D. degree in electrical engineering from the Technical University of Catalonia (UPC), Barcelona, Spain, in 1987 and 1991, respectively. From 1987 to 1994, he was a Researcher with UPC, where he became an Associate Professor, in 1992. In 1995, he joined the Department of Mathematics and Informatics, University of the Balearic Islands (UIB), Mallorca, Spain, where he became Full Professor, in 2010. He is currently leading the Mobile Communications Group at UIB, where he has been the Project Manager of projects ARAMIS, DREAMS, DARWIN, MARIMBA, COSMOS, ELISA, and TERESA, all of them funded by the Spanish and Balearic Islands Governments. In the past, he was also involved with several European projects (ATDMA, CODIT, and COST). His current research interests and activities span the fields of digital communications theory and wireless communication systems, with particular emphasis on radio resource management strategies applied to 5G and 6G wireless networks. On these topics, he has published more than 100 journal and conference papers, and some book chapters. Dr. Femenias was a recipient of the Best Paper Awards at the 2007 IFIP International Conference on Personal Wireless Communications and at the 2009 IEEE Vehicular Technology Conference - Spring. He has served for various IEEE conferences as a Technical Program Committee Member, as the Publications Chair for the IEEE 69th Vehicular Technology Conference (VTC-Spring 2009) and as a Local Organizing Committee member of the IEEE Statistical Signal Processing (SSP 2016).

NARJES LASSOUED successfully completed the preparatory cycle in mathematics and physics, in 2009. She received the Diploma of Engineer degree in communications and networking from the National Engineering School of Gabes (ENIG), Tunisia, in 2012, and the Diploma of Master degree in communication systems from the National Engineering School of Tunis (ENIT), Tunisia. She is currently pursuing the Ph.D. degree with the National Engineering School of Gabes (ENIG), Tunisia. She is also with the University of Carthage, Gabes, Tunisia. She is also a Research Member of the Research Unit of Innovation of Communicating and Cooperative Mobile (innov'COM), Superior School of Communications of Tunis (Sup'Com).

FELIP RIERA-PALOU (Senior Member, IEEE) received the dual B.S./M.S. degrees in computer engineering from the University of the Balearic Islands (UIB), Mallorca, Spain, in 1997, the M.Sc. and Ph.D. degrees in communication engineering from the University of Bradford, U.K., in 1998 and 2002, respectively, and the M.Sc. degree in statistics from the University of Sheffield, U.K., in 2006. From May 2002 to March 2005, he was with Philips Research Laboratories (Eindhoven, The Netherlands) first as a Marie Curie Postdoctoral Fellow (European Union) and later as a Member of technical staff. While at Philips he worked on research programs related to wideband speech/audio compression and speech enhancement for mobile telephony. From April 2005 to December 2009, he was a Research Associate (Ramon y Cajal program, Spanish Ministry of Science) with the Mobile Communications Group, Department of Mathematics and Informatics, UIB. Since January 2010, he has been an Associate Research Professor (I3 program, Spanish Ministry of Education) with UIB. His current research interests are in the general areas of signal processing and wireless communications.