

Joint Power Allocation and Access Point Selection for Cell-free Massive MIMO

Thang X. Vu*, Symeon Chatzinotas*, Shahram ShahbazPanahi*[†], and Bjorn Ottersten*

* The Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg

[†] Faculty of Engineering and Applied Science, University of Ontario Institute of Technology, Oshawa, Canada.

Email: {thang.vu, symeon.chatzinotas, shahram.shahbazpanahi, bjorn.ottersten}@uni.lu

Abstract—Cell-free massive multiple-input multiple-output (CF-MIMO) is a promising technological enabler for fifth generation (5G) networks in which a large number of access points (APs) jointly serve the users. Each AP applies conjugate beamforming to precode data, which is based only on the AP's local channel state information. However, by having the nature of a (very) large number of APs, the operation of CF-MIMO can be energy-inefficient. In this paper, we investigate the energy efficiency performance of CF-MIMO by considering a practical energy consumption model which includes both the signal transmit energy as well as the static energy consumed by hardware components. In particular, a joint power allocation and AP selection design is proposed to minimize the total energy consumption subject to given quality of service (QoS) constraints. In order to deal with the combinatorial complexity of the formulated problem, we employ norm $l_{2,1}$ -based block-sparsity and successive convex optimization to leverage the AP selection process. Numerical results show significant energy savings obtained by the proposed design, compared to all-active APs scheme and the large-scale based AP selection.

Index terms— Free-cell massive MIMO, energy minimization, block sparsity, difference of convex, successive convex approximation, iterative optimization.

I. INTRODUCTION

Cell-free massive multiple-input multiple-output (CF-MIMO) is a promising technology to further improve the wireless networks in terms of energy efficiency and coverage. Synchronous time-division duplex (TDD) CF-MIMO is operated based on the cooperation of a very large number of access points (APs) in the same frequency-time resource block without cell-barrier, unlike conventional cell-based wireless networks. As a result, CF-MIMO can exploit the favourable propagation and channel hardening properties similar to conventional massive MIMO to efficiently suppress inter-user interference, thereby resulting in improved spectral efficiency [1]. Another favourable property of CF-MIMO is simplicity of signal precessing at the APs as the APs only require local channel state information (CSI) and apply low-complexity conjugate precoding design, resulting in great scalability and relatively low overhead among the APs [2]. The potential of

CF-MIMO can not be harvested if the power allocation is not properly designed. The authors of [3] propose a low-complexity power control based on zero-forcing precoding to improve the system energy efficiency (EE) of the CF-MIMO downlink with single-antenna APs. The extension of this approach for multiple-antenna APs is studied in [2]. In [4], the authors propose a power allocation technique to maximize the downlink global EE in millimeter wave CF-MIMO. The design of CF-MIMO uplinks are investigated in [5–7], in which the symbol transmit power and receiver filter coefficients are jointly optimized. It is worth noting that these studies assume all APs are active and focus on power allocation.

In practice, the energy consumption of an AP consists of not only the energy of electromagnetic radiation but also the energy consumed in hardware components, e.g., radio frequency (RF) chains, as well as the energy consumed in the backhaul. If an AP has negligible contribution to the system performance, it is beneficial to turn off that AP to save energy. This motivates us to investigate the joint design of AP activation and power allocation to further improve the energy usage in CF-MIMO. Large-scale based AP selection has been proposed [2], [8] in which each user is served by only some APs in the proximity. This method, however, does not guarantee the same selection patterns for all users, hence an AP may still be active even when it serves only a single user. It is worth noting that the selection policy in [9] is not applicable to CF-MIMO schemes as this policy was designed for co-located massive MIMO with the same pathloss from the co-located antennas to each user, whereas the pathloss from the APs to a user are different in CF-MIMO. Moreover, [9] only estimates the number of active antennas without specifying them, which is not sufficient in CF-MIMO schemes.

In this paper, we investigate the energy efficiency performance of TDD CF-MIMO. In particular, we propose joint power allocation and AP selection technique based on the block-sparsity norm $l_{2,1}$ to guarantee the same AP selection pattern for all users. In order to overcome the non-convexity

of the formulated problem, we propose an iterative algorithm which solves the reformulated problem based on successive convex approximation. Finally, the effectiveness of the proposed joint design is demonstrated via numerical results, which show a significant energy reduction compared to the large-scale based AP selection [2], [8] and all-active APs scheme.

The rest of the paper is organised as follows. Section II presents the system and signal transmission models. Section III proposes the joint AP selection and power allocation design. Section IV shows numerical results. Finally, Section V concludes the paper.

II. SYSTEM AND SIGNAL TRANSMISSION MODEL

We consider a cell-free massive MIMO communication system in which N access points (APs) coherently serve K users that are randomly distributed on the coverage area. Each AP is equipped with M antennas, while the users are assumed to have a single-antenna. A central processing unit (CP) connects with the APs via a backhaul network.

We consider quasi-static block Rayleigh fading channels, in which the channel gain between the APs and the users are assumed to be constant within a transmission frame and independently change from one frame to another, and is mutually independent across the links. The channel coefficient includes a slow fading component, e.g., pathloss and shadowing, and a fast fading component. The channel fading coefficient between AP n and user k is modeled as $\mathbf{g}_{nk} = \beta_{nk}^{1/2} \mathbf{h}_{nk}$, where β_{nk} represents the pathloss, which is assumed to be constant over a very large number of frames, and $\mathbf{h}_{nk} \in \mathbb{C}^{M \times 1}$ is the fast fading part following circular-symmetric complex Gaussian distribution $\mathbf{h}_{nk} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$, with \mathbf{I}_M being an identity matrix of size $M \times M$.

A. Uplink pilot-based channel estimation

Before transmitting data, each AP acquires CSI knowledge via uplink pilot-assisted channel estimation. The transmission is performed every block of length τ_c channel uses (c.u.), which must be smaller than the channel coherence time. Each transmission frame is divided into two phases: uplink training and downlink data transmission. No downlink pilot is needed as the users use the first-order channel statistic for decoding [2]. In the training phase, the users send (orthogonal) known pilot sequences to the APs simultaneously, through which the APs obtain the CSI and then use the so-obtained CSI for data transmission, by assuming the channel reciprocity. As for the TDD, $\tau_p < \tau_c$ c.u. are used for training. Let $\varphi_k \in \mathbb{C}^{\tau_p \times 1}$ be a pilot signal dedicated to user k , where $\|\varphi_k\|^2 = 1, \forall k$. Assuming synchronization across the network, the $M \times \tau_p$ received signals matrix corresponding to pilot transmission at the n -th AP is given as

$$\mathbf{Y}_{p,n} = \sqrt{\rho} \sum_{k=1}^K \mathbf{g}_{nk} \varphi_k^H + \mathbf{Z}_n, \quad (1)$$

where ρ is the pilot's transmit power, $\mathbf{Z}_n \in \mathbb{C}^{M \times \tau_p}$ is the Gaussian noise matrix whose elements follow $\mathcal{CN}(0, \sigma^2)$ with σ^2 being the noise power.

The APs apply minimum mean square error (MMSE) estimation to obtain the estimate of \mathbf{g}_{nk} , denoted as $\hat{\mathbf{g}}_{nk}$. The MMSE channel estimate is given as

$$\hat{\mathbf{g}}_{nk} = \mathbb{E}\{\mathbf{g}_{nk} \tilde{\mathbf{y}}_{nk}^H\} (\mathbb{E}\{\tilde{\mathbf{y}}_{nk} \tilde{\mathbf{y}}_{nk}^H\})^{-1} \tilde{\mathbf{y}}_{nk}, \quad (2)$$

where $\tilde{\mathbf{y}}_{nk} = \mathbf{Y}_{p,n} \varphi_k = \sqrt{\rho} \mathbf{g}_{nk} + \sqrt{\rho} \sum_{j \neq k} \mathbf{g}_{nj} \varphi_j^H \varphi_k + \mathbf{Z}_n \varphi_k$. From (1) and (2), we obtain

$$\hat{\mathbf{g}}_{nk} = \frac{\sqrt{\rho} \beta_{nk}}{\rho \sum_{j=1}^K \beta_{nj} |\varphi_j^H \varphi_k|^2 + \sigma^2} \tilde{\mathbf{y}}_{nk}. \quad (3)$$

It is noted that the channel estimate $\hat{\mathbf{g}}_{nk}$ consists of M independent elements. The variance of the m -th element of $\hat{\mathbf{g}}_{nk}$ is given as $\gamma_{nk} = \mathbb{E}\{|\hat{\mathbf{g}}_{nk}|^2\} = \frac{\rho \beta_{nk}^2}{\rho \sum_{j=1}^K \beta_{nj} |\varphi_j^H \varphi_k|^2 + \sigma^2}$. The MMSE estimation error is $\tilde{\mathbf{g}}_{nk} = \mathbf{g}_{nk} - \hat{\mathbf{g}}_{nk}$, which is independent from the true channel. In addition, each element of the channel estimate error follows $\mathcal{CN}(0, \beta_{nk} - \gamma_{nk})$.

Remark 1: If $\tau_p \geq K$, one can guarantee mutually orthogonal pilots among the K users, i.e., $\varphi_i^T \varphi_k = \delta_{ik}$, where δ_{ik} is the delta function. In this case, for a given total training power ρ , the CSI accuracy in (3) does not depend on the pilot length.

B. Downlink data transmission

Upon acquiring the CSI by the APs, the CP sends the data symbols to the APs, which then precode the data before transmitting to the users. Due to large-scale volume of the networks, a low-complexity precoding method is required. Therefore, we adopt conjugate precoding design, which offers a low-complexity level since it is done locally and no CSI exchange is required among the APs. In addition, it is shown that as the number of APs is very large, this precoding method achieves a close performance to the centralized method [10].

Let c_k denote a data symbol dedicated for user k . After receiving all the data symbols, AP n obtains the $M \times 1$ vector \mathbf{x}_n by using the conjugate precoding, as follows:

$$\mathbf{x}_n = \sum_{k=1}^K \sqrt{p_{nk}} \hat{\mathbf{g}}_{nk}^* c_k, \quad (4)$$

where p_{nk} is the power scaling factor designed for user k from AP n . Limited by the maximum transmit power P_{TX} at the AP, we have $\mathbb{E}\{\|\mathbf{x}_n\|^2\} = M \sum_{k=1}^K p_{nk} \gamma_{nk} \leq P_{TX}$.

The received signal at user k is given as

$$\begin{aligned} y_k &= \sum_{n=1}^N \mathbf{g}_{nk}^T \mathbf{x}_n + z_k = \sum_{n=1}^N \sum_{k=1}^K \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^* c_k + z_k \\ &= \sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^* c_k + \sum_{n=1}^N \sum_{k \neq j=1}^K \sqrt{p_{nj}} \mathbf{g}_{nj}^T \hat{\mathbf{g}}_{nj}^* c_j + z_k. \end{aligned} \quad (5)$$

Due to lack of downlink training, user k does not have knowledge about the effective channel gain $\sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^*$. It instead employs the first-order channel statistic, i.e., $\mathbb{E}\{\sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^*\}$, as the estimated channel. The received signal can be rewritten as

$$y_k = \mathbb{E}\left\{ \sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^* \right\} c_k$$

$$\begin{aligned}
& + \left(\sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^* - \mathbb{E} \left\{ \sum_{n=1}^N \sqrt{p_{nk}} \mathbf{g}_{nk}^T \hat{\mathbf{g}}_{nk}^* \right\} \right) c_k \\
& + \sum_{n=1}^N \sum_{k \neq j=1}^K \sqrt{p_{nj}} \mathbf{g}_{nj}^T \hat{\mathbf{g}}_{nj}^* c_j + z_k. \quad (6)
\end{aligned}$$

By using the capacity bound and taking the pilot duration into account, the effective achievable rate of user k is given as [2]:

$$\begin{aligned}
R_k &= \frac{\tau_c - \tau_p}{\tau_c} \times \\
& B \log_2 \left(1 + \frac{M^2 \mathbf{p}_k^T \mathbf{\Gamma}_{kk} \mathbf{p}_k}{M^2 \sum_{j \neq k} \mathbf{p}_j^T \mathbf{\Gamma}_{jk} \mathbf{p}_j + M \sum_{j=1}^K \mathbf{p}_j^T \mathbf{D}_{jk} \mathbf{p}_j + \sigma^2} \right), \quad (7)
\end{aligned}$$

where B is the channel bandwidth, $\mathbf{p}_k \triangleq [\sqrt{p_{1k}}, \sqrt{p_{2k}}, \dots, \sqrt{p_{Nk}}]^T$, $\mathbf{D}_{jk} \in \mathbb{R}^{N \times N}$ is the diagonal matrix whose n -th diagonal element equals to $\gamma_{nj} \beta_{nk}$, and $\mathbf{\Gamma}_{jk} = \bar{\gamma}_{jk} \bar{\gamma}_{jk}^T$ with $\bar{\gamma}_{jk} = |\boldsymbol{\varphi}_j^H \boldsymbol{\varphi}_k| \left[\gamma_{1j} \frac{\beta_{1k}}{\beta_{1j}}, \dots, \gamma_{Nj} \frac{\beta_{Nk}}{\beta_{Nj}} \right]^T$.

The sum-rate of all users is given by

$$R_{\Sigma} = \sum_{k=1}^K R_k. \quad (8)$$

C. Energy consumption model

We consider practical energy consumption model which takes into account both electromagnetic radiation and hardware-related static power, e.g., RF chain and amplifier. Let P_{tot} denote the total energy consumption for serving the users, which is calculated as

$$P_{tot} = \sum_{n=1}^N P_{A,n} + \sum_{n=1}^N P_{B,n},$$

where $P_{A,n}$ is the power consumed by AP n and $P_{B,n}$ is the power consumption in the backhaul associated to AP n . The transmit power at the APs consists of two parts: circuit power P_{ap} , e.g., the power consumption of the RF chains and amplifiers, and the emitting power. Hence, we can write

$$P_{A,n} = P_{ap} + \sum_{k=1}^K \gamma_{nk} p_{nk}.$$

The backhaul capacity is assumed to be sufficient to support the data rate. The backhaul power consumption consists of two parts: P_{fix} which is independent from backhaul traffic and P_{bh} - the traffic-dependent power (in W/bps) consumed in the n -th AP's backhaul. Hence we can write $P_{B,n} = P_{fix} + P_{bh} R_{\Sigma}$.

The total power consumption can be written as:

$$P_{tot} = N(P_{fix} + P_{ap}) + NP_{bh} R_{\Sigma} + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} p_{nk}. \quad (9)$$

III. JOINT AP SELECTION AND POWER ALLOCATION

We aim to minimize the total energy consumption of CF-MIMO system via the joint power allocation and AP activation design. Different from other studies, which focus only on transmit power, our work considers energy consumption on both backhaul links and RF chains. We note that the total power consumption in (9) stands for the case when all the

APs are active. When only a subset $\mathcal{A} \subset \mathcal{N} \triangleq \{1, 2, \dots, N\}$ of the APs are selected, the total energy consumption becomes

$$P_{tot, \mathcal{A}} = NP_{fix} + |\mathcal{A}|(P_{ap} + P_{bh} R_{\Sigma}) + \sum_{n \in \mathcal{A}} \sum_{k=1}^K \gamma_{nk} p_{nk}, \quad (10)$$

where $|\mathcal{A}|$ denotes the cardinality of the set \mathcal{A} .

Since the first term in $P_{tot, \mathcal{A}}$ does not depend on \mathcal{A} and $\{p_{nk}\}$, it can be excluded when minimizing the total power consumption, without loss of optimality. The joint optimization problem is then formulated as follows:

$$\text{Minimize}_{p_{nk} \geq 0, \mathcal{A} \subset \mathcal{N}} |\mathcal{A}| P_{ap} + |\mathcal{A}| P_{bh} \sum_{k=1}^K R_k + \sum_{n \in \mathcal{A}} \sum_{k=1}^K \gamma_{nk} p_{nk} \quad (11)$$

$$\text{s.t. } R_k \geq \eta_k, \forall k, \quad (11a)$$

$$\sum_{k=1}^K \gamma_{nk} p_{nk} \leq P_{TX}, \forall n \in \mathcal{A} \quad (11b)$$

where R_k is given in (7), \mathcal{A} is the subset of the activated APs, η_k is the QoS requirement at user k , and P_{TX} is the maximum transmit power at each AP.

The difficulty in solving (11) lies not only in the non-convexity of the effective rate R_k but also in finding the set \mathcal{A} of the selected APs. To add insult to injury, the second term in the objective function makes it even more challenging. Intuitively, one can solve (11) by finding the optimal power allocation for every \mathcal{A} (though the resulting problem is still non-convex). This method, however imposes combinatorial complexity since there are $\sum_{n=\lfloor K/M \rfloor + 1}^N \binom{N}{n}$ possible subsets in total. For example, for the CF-MIMO with $N = 30$ APs and $K = 10$ users, the total number of antenna subsets exceeds one billion, which makes the problem practically infeasible.

To overcome this challenge, we propose to optimize the transmit power and activate the APs simultaneously. The main idea behind the proposed solution is that the activated APs can be inherently captured by the power allocation factors, if they are properly designed [11], [12]. Before proceeding, without loss of generality, we use new (power allocation) variable $q_{nk} \triangleq \sqrt{p_{nk}}$ for ease of presentation. Denote $\mathbf{q}_k = [q_{1k}, q_{2k}, \dots, q_{Nk}]^T$ and $\bar{\mathbf{q}}_n \triangleq [q_{n1}, q_{n2}, \dots, q_{nK}]$. Furthermore, let $\bar{\mathbf{Q}} = [\|\bar{\mathbf{q}}_1\|_2, \dots, \|\bar{\mathbf{q}}_N\|_2]^T$. Obviously, $\|\bar{\mathbf{Q}}\|_0$ represents the number of active APs, i.e., AP n is active iff $\prod_k q_{nk} \neq 0$. Employing $\|\bar{\mathbf{Q}}\|_0$, we can reformulate problem (11) as follows:

$$\text{Minimize}_{q_{nk} \geq 0} \|\bar{\mathbf{Q}}\|_0 (P_{ap} + P_{bh} \sum_{k=1}^K \bar{R}_k) + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} p_{nk} \quad (12)$$

$$\text{s.t. } \bar{R}_k \geq \eta_k, \forall k, \quad (12a)$$

$$\sum_{k=1}^K \gamma_{nk} q_{nk}^2 \leq P_{TX}, \forall n, \quad (12b)$$

where \bar{R}_k is a new presentation of R_k with respect to q_{nk} and defined as $\bar{R}_k \triangleq \frac{\tau_c - \tau_p}{\tau_c} B \log_2 \left(1 + \frac{M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k}{M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2} \right)$.

By introducing auxiliary positive variables $u, x_k, y_k, z_k, \forall k$,

we can equivalently reformulate problem (12) as

$$\text{Minimize}_{q_{nk}, x_k, y_k, z_k, u} P_{\text{ap}}u + \bar{B}P_{\text{bh}} \sum_{k=1}^K x_k + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} q_{nk}^2 \quad (13)$$

$$\text{s.t.} \quad \frac{M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k}{M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2} \geq \bar{\eta}_k, \forall k \quad (13a)$$

$$u y_k \leq x_k, \forall k \quad (13b)$$

$$\|\bar{\mathbf{Q}}\|_0 \leq u \quad (13c)$$

$$\log(1 + z_k) \leq y_k, \forall k \quad (13d)$$

$$\frac{M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k}{M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2} \leq z_k, \forall k, \quad (13e)$$

where $\bar{\eta}_k = 2^{\eta_k/\bar{B}} - 1$ and $\bar{B} = \frac{\tau_c - \tau_p}{\tau_c} B$.

The equivalence between (13) and (12) can be directly shown at the optimum since equalities hold in all constraints. Unfortunately, problem (13) is still difficult due to the norm-zero constraint in (13c) and non-convexity of the constraints (13a), (13b), (13d) and (13e). To deal with the former, one can relax the norm-zero with l_1 -norm to convexify (13c). This method, however, might result in different AP selection patterns for different users, which eventually fails to properly select the APs.

In this paper, we employ block-sparsity norm $l_{2,1}$ to guarantee same AP selection patterns for all the users in (13). The norm $l_{2,1}$ of matrix \mathbf{X} of size $N \times M$ is defined as

$$\|\mathbf{X}\|_{1,2} = \sum_{n=1}^N \sqrt{\sum_{m=1}^M X_{n,m}^2}.$$

Based on the norm $l_{2,1}$, constraint (13c) can be approximated by $\|\mathbf{Q}\|_{2,1} \leq \lambda u$, where $\mathbf{Q} \triangleq [q_{nk}]_{N \times K}$ and λ is the scaling factor. Then we can reformulate problem (13) as follows:

$$\text{Minimize}_{q_{nk}, x_k, y_k, z_k, u} P_{\text{ap}}u + \bar{B}P_{\text{bh}} \sum_{k=1}^K x_k + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} q_{nk}^2 \quad (14)$$

$$\text{s.t.} \quad M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2 \leq \frac{1}{\bar{\eta}_k} M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k, \forall k \quad (14a)$$

$$(u + y_k)^2 \leq 2x_k + u^2 + y_k^2, \forall k \quad (14b)$$

$$\|\mathbf{Q}\|_{2,1} \leq \lambda u \quad (14c)$$

$$1 + z_k \leq e^{y_k}, \forall k \quad (14d)$$

$$\frac{M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k}{z_k} \leq M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2, \forall k. \quad (14e)$$

In (14), the first constraint is obtained from (13a) since all the elements are positive. Constraint (14b) is obtained by adding $u^2 + y_k^2$ to both sides of (13b). Constraint (14d) is equivalent to (13d) since the exponential function is monotonic. Finally, constraint (14e) is equivalent to (13e) because both z_k and q_k

TABLE I: ITERATIVE ALGORITHM TO SOLVE (19)

-
1. Initialize a feasible solution $\hat{\mathbf{q}}_k, \hat{u}, \hat{y}_k$.
 $\epsilon, t = 1, E_{\text{old}}, t_{\text{MAX}}$ and $\text{error} = 1$.
 2. While $\text{error} > \epsilon$ and $t < t_{\text{MAX}}$ do
 - 2.1. Solve the problem (19) to obtain $\mathbf{p}_k^*, u^*, x_k^*, y_k^*, z_k^*, \forall k$. Compute $E^{(t)} = P_{\text{ap}}u^* + \bar{B}P_{\text{bh}} \sum_{k=1}^K x_k^* + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} (q_{nk}^*)^2$
 - 2.3. Compute $\text{error} = |E^{(t)} - E_{\text{old}}|$
 - 2.4. Update $\hat{\mathbf{q}}_k = \mathbf{q}_k^*; \hat{u} = u^*; \hat{y}_k = y_k^*$,
 $E_{\text{old}} = E^{(t)}, t := t + 1$.
-

are positive.

Although the objective function and constraint (14c) are convex, solving problem (14) is still challenging due to the non-convexity of all other constraints. Fortunately, since function $\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{y}$ is jointly convex in \mathbf{x} and $y > 0$ (see Appendix A) and the right-hand side of these constraints are convex, one can employ successive approximation method to solve (14) in an iterative manner. The idea behind the successive approximation is to use the first-order approximation in the right-hand-side (RHS) of the constraints and refine the approximations iteratively.

In particular, denote $\hat{u}, \hat{\mathbf{q}}_k, \hat{y}_k, \forall k$ as the solution of the approximated problem in iteration t . In the $(t+1)$ -th iteration, we approximate the RHS of (14a) by its first-order approximation at $\hat{\mathbf{q}}_k$. As a result, the constraint (14a) is approximated by

$$M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2 \leq \frac{M^2}{\bar{\eta}_k} \left[\mathbf{q}_k^T (\mathbf{\Gamma}_{kk}^T + \mathbf{\Gamma}_{kk}) \hat{\mathbf{q}}_k - \hat{\mathbf{q}}_k^T \mathbf{\Gamma}_{kk}^T \hat{\mathbf{q}}_k \right]. \quad (15)$$

With the help of the first-order approximation of $\mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j$ and $\mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j$ in the RHS of (14e), and noting that \mathbf{D}_{jk} is symmetric, we can approximate (14e) by

$$\begin{aligned} \frac{M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k}{z_k} &\leq M^2 \sum_{j \neq k} \mathbf{q}_j^T (\mathbf{\Gamma}_{jk}^T + \mathbf{\Gamma}_{jk}) \hat{\mathbf{q}}_j \\ &+ 2M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \hat{\mathbf{q}}_j - M^2 \sum_{j \neq k} \hat{\mathbf{q}}_j^T \mathbf{\Gamma}_{jk}^T \hat{\mathbf{q}}_j \\ &- M \sum_{j=1}^K \hat{\mathbf{q}}_j^T \mathbf{D}_{jk} \hat{\mathbf{q}}_j + \sigma^2. \end{aligned} \quad (16)$$

Similarly, we can approximate constraint (14b) by

$$(u + y_k)^2 \leq 2x_k + \hat{u}(2u - \hat{u}) + \hat{y}_k(2y_k - \hat{y}_k) \quad (17)$$

and approximate (14d) by

$$1 + z_k \leq e^{\hat{y}_k} (y_k - \hat{y}_k + 1). \quad (18)$$

Table I provides detailed steps of the proposed iterative algorithm to solve (14).

$$\begin{aligned} & \underset{q_{nk}, x_k, y_k, z_k, u}{\text{Minimize}} && P_{\text{ap}}u + \bar{B}P_{\text{bh}} \sum_{k=1}^K x_k + \sum_{n=1}^N \sum_{k=1}^K \gamma_{nk} q_{nk}^2 \quad (19) \\ & \text{s.t.} && (14\text{c}), (15), (16), (17), (18). \end{aligned}$$

Proposition 1 (Convergence): The objective function of (19) decreases over the iterations and the proposed iterative algorithm in Table I converges at least to local optimum. The proof of Proposition 1 can be found by using a technique similar to the one presented in [13]. Although Proposition 1 does not prove the global optimality, it gives the justification for the iterative algorithm.

Remark 2 (Initialization): The convergence of problem (19) closely depends on the initial values in the iterative algorithm. Proper initial values not only makes the algorithm feasible but also speeds up the convergence. In order to obtain the initial value $\hat{q}_k, \hat{u}, \hat{y}_k$, we first run the following feasible problem:

$$\begin{aligned} & \underset{q_{nk}}{\text{Minimize}} && 1 \quad (20) \\ & \text{s.t.} && M^2 \sum_{j \neq k} \mathbf{q}_j^T \mathbf{\Gamma}_{jk} \mathbf{q}_j + M \sum_{j=1}^K \mathbf{q}_j^T \mathbf{D}_{jk} \mathbf{q}_j + \sigma^2 \\ & && \leq \frac{1}{\bar{\eta}_k} M^2 \mathbf{q}_k^T \mathbf{\Gamma}_{kk} \mathbf{q}_k, \forall k \\ & && \sum_k \gamma_{nk} q_{nk}^2 \leq P_{TX}, \forall n. \end{aligned}$$

Let \mathbf{q}_k^* be a feasible solution to (20), we can calculate the initial values of the proposed iterative algorithm as

$$\hat{q}_k = \mathbf{q}_k^*; \quad \hat{u} = \|\mathbf{Q}^*\|_{2,1}/\lambda,$$

and $\hat{y}_k =$

$$\log_2 \left(1 + \frac{M^2 (\mathbf{q}_k^*)^T \mathbf{\Gamma}_{kk} \mathbf{q}_k^*}{M^2 \sum_{j \neq k} (\mathbf{q}_j^*)^T \mathbf{\Gamma}_{jk} \mathbf{q}_j^* + M \sum_{j=1}^K (\mathbf{q}_j^*)^T \mathbf{D}_{jk} \mathbf{q}_j^* + \sigma^2} \right).$$

IV. NUMERICAL RESULTS

In this section, we demonstrate the effectiveness of the proposed joint design via numerical results. The APs are located on a uniform grid in the area of 1 km \times 1 km at the height of 15 meters, while the users are randomly distributed in this area. The system operates at 2.4 GHz with 20 MHz bandwidth. The path loss is followed by the WINNER II b5f [14], in which the attenuation in dB is given as $PL(d) = 57.5 + 23.5 \log(d) + 23 \log(f/5)$, where d is the distance in meter and f is the operating frequency in GHz. The noise power is -110 dBW. Each transmission block spreads over $\tau_c = 1000$ c.u., in which $\tau_p = K$ c.u. is used for channel estimation. Other parameters are in Table II. The proposed joint design is compared with to i) the so-called *All APs*, in which all APs are active and ii) *Large-scale based* AP selection [2], [8].

Fig. 1 compares the total power consumption of the proposed joint design with the two reference schemes for different number of users. For serving the same user requests, the proposed joint design significantly reduces the energy consumption, compared to the two references. Specifically, the

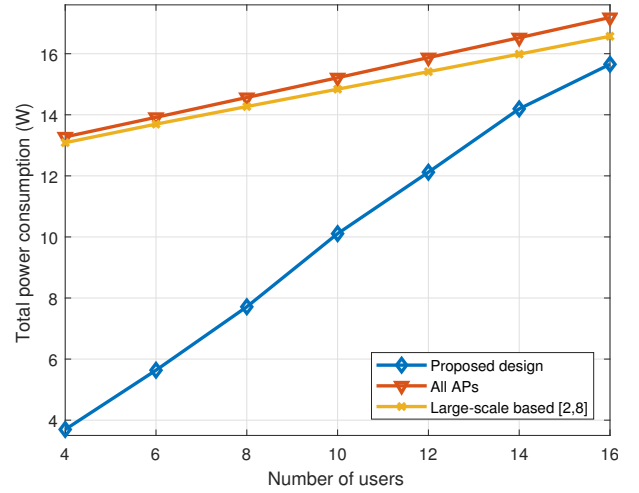


Fig. 1: Total power consumptions for different number of users K . All users require a minimum QoS of 20 Mbps. $P_{\text{bh}} = 0.5$ W/Gbps.

joint design consumes 55% and 33% less energy than two reference schemes for $K = 6$ and $K = 10$, respectively. This is because in this case, the joint design only uses the "best" APs to serve the users, compared to all APs in first reference. It is also shown that the *Large-scale based* selection scheme [2], [8] improves total energy usage compared with the *All APs* but consumes much more than the joint design. This is because the *Large-scale based* scheme does not guarantee the same AP selection patterns for the users, thus, the AP is active even when it serves only one user. When there are more users in the network, the joint design's power consumption approaches the references, as it requires more APs to efficiently mitigate inter-user interference.

Fig. 2 compares the total energy consumption as a function of the traffic-dependent cost on backhaul, i.e., P_{bh} . It is observed that the total energy consumption of all schemes is approximately linearly dependent on P_{bh} , which can be seen from the energy consumption model in (9). In all cases, the joint design spends much less energy, compared to the reference schemes.

Fig. 3 plots the total energy consumption versus the same QoS requirements $\eta = \eta_k, \forall k$. For small QoS requirements, e.g., 20 Mbps, all users can be served when only a subset of the APs are activated. This is revealed in Fig. 3 via significant power saving offered by the proposed joint design

TABLE II: Simulation parameters

Parameters	Value
Number of APs N	30
Number of antennas per AP M	2
Number of users K	4 to 16
Fix power consumption/backhaul P_{fix}	0.01 W
Circuit power consumption/antenna P_{ap}	0.2 W
Traffic-dependent power/Gbps P_{bh}	0.1 - 0.9 W

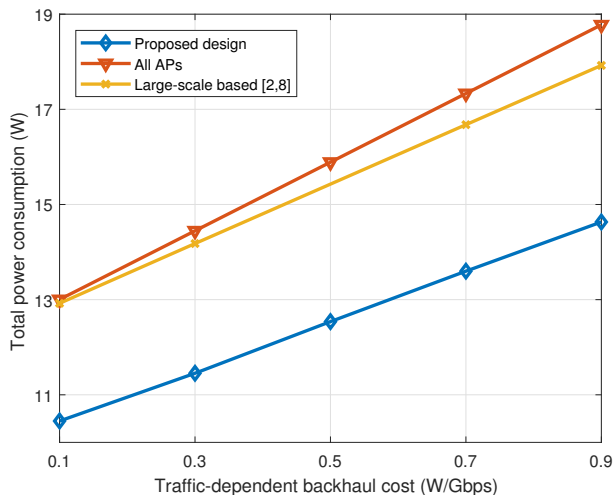


Fig. 2: Total power consumptions versus the traffic-dependent power consumption P_{bh} on the backhaul. $K = 10$, $\eta_k = 24\text{Mbps}$, $\forall k$.

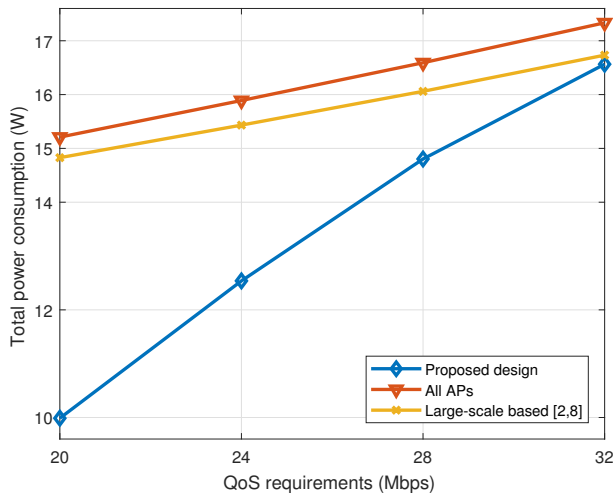


Fig. 3: Total power consumptions as a function of the minimum rate requirement η . $K = 10$, $P_{\text{bh}} = 0.1$ (W/Gbps).

problem, compared to the two references. Having higher QoS requirements results in smaller energy saving of the joint design, as it requires more APs to serve the users.

V. CONCLUSIONS

We have investigated the performance of cell-free massive MIMO systems under realistic energy consumption model which takes into account both electromagnetic radiation and energy consumed by hardware components as well as the energy consumed by the backhaul. We proposed a joint design for AP activation and power allocation based on block-sparsity norm $l_{2,1}$ to minimize the total power consumption while satisfying the predefined QoS requirements as well as per-AP transmit power constraint. It is shown that the proposed design significantly reduces the total energy consumption for moderate network loads, compared with the all-activated AP and the large-scale based selection. The proposed design

suggests a promising dynamic management framework for cell-free massive MIMO to optimize other metrics, e.g., sum-rate or energy efficiency.

ACKNOWLEDGEMENT

This work is supported by the projects FNR CORE ProCAST, ELECTIC, PROSAT, and DISBUS.

APPENDIX A

CONVEXITY OF FUNCTION $\frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{y}$

To prove the convexity of $F(\mathbf{x}, y) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{y}$ for any positive semi-definite metric \mathbf{A} , we need to show that the Hessian metric of $F(\mathbf{x}, y)$ is positive semidefinite. Indeed, the Hessian matrix of $F(\mathbf{x}, y)$ is

$$\mathbf{H}_F = \begin{bmatrix} \frac{\mathbf{A} + \mathbf{A}^T}{y} & -\frac{(\mathbf{A} + \mathbf{A}^T)\mathbf{x}}{y^2} \\ -\frac{\mathbf{x}^T(\mathbf{A} + \mathbf{A}^T)}{y^2} & \frac{2\mathbf{x}^T \mathbf{A} \mathbf{x}}{y^3} \end{bmatrix}.$$

For arbitrary vector $\mathbf{c} = [\mathbf{a}^T b]^T$, where $\mathbf{a} \in \mathbb{R}^{N \times 1}$, consider a function

$$\begin{aligned} \mathbf{c}^T \mathbf{H}_F \mathbf{c} &= \frac{\mathbf{a}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{a}}{y} - \frac{\mathbf{a}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{x} b}{y^2} \\ &\quad - \frac{\mathbf{x}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{a} b}{y^2} + \frac{2\mathbf{x}^T \mathbf{A} \mathbf{x} b^2}{y^3} \\ &\stackrel{(*)}{=} \frac{\mathbf{a}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{a}}{y} - 2 \frac{\mathbf{a}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{x} b}{y^2} + \frac{\mathbf{x}^T (\mathbf{A} + \mathbf{A}^T) \mathbf{x} b^2}{y^3} \\ &= \frac{\mathbf{a}^T \tilde{\mathbf{A}} \mathbf{a} - 2\mathbf{a}^T \tilde{\mathbf{A}} \tilde{\mathbf{x}} + \tilde{\mathbf{x}}^T \tilde{\mathbf{A}} \tilde{\mathbf{x}}}{y}, \end{aligned} \quad (21)$$

where $\tilde{\mathbf{A}} \triangleq \mathbf{A}^T + \mathbf{A}$, $\tilde{\mathbf{x}} \triangleq \mathbf{x}b/y$ and $(*)$ results from the fact that \mathbf{A} is symmetric and $\mathbf{a}^T \tilde{\mathbf{A}} \tilde{\mathbf{x}} = \mathbf{x}^T \tilde{\mathbf{A}} \mathbf{a}$. It is obvious that the RHS of (21) is always non-negative for $y > 0$ and positive semi-definite matrix $\tilde{\mathbf{A}}$, which concludes the positive semi-definite of the Hessian metric of $F(\mathbf{x}, y)$.

REFERENCES

- [1] 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.
- [2] H. Q. Ngo, L. 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.
- [3] 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.
- [4] M. Alonzo, S. Buzzi, A. Zappone, and C. Dielma, "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.
- [5] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink max-min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.
- [6] Y. Zhang, H. Cao, M. Zhou, and L. Yang, "Power optimization for energy efficiency in cell-free massive MIMO with ZF receiver," in *Proc. Int. Conf. Adv. Commun. Technol.*, Feb. 2019, pp. 366–371.
- [7] M. Bashar, K. Cumanan, A. G. Burr, H. Q. Ngo, E. G. Larsson, and P. Xiao, "On the energy efficiency of limited-backhaul cell-free massive MIMO," in *Proc. IEEE Int. Conf. Commun.*, May 2019, pp. 1–7.
- [8] S. Buzzi and C. DiAndrea, "Cell-free massive MIMO: User-centric approach," *IEEE Wireless Commun. Lett.*, vol. 6, no. 6, pp. 706–709, Dec. 2017.

- [9] G. Dong, H. Zhang, S. Jin, and D. Yuan, "Energy-efficiency-oriented joint user association and power allocation in distributed massive MIMO systems," *IEEE Trans. Veh. Technol.*, vol. 68, no. 6, pp. 5794–5808, June 2019.
- [10] 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, Aug. 2016.
- [11] O. Mehanna, N. D. Sidiropoulos, and G. B. Giannakis, "Joint multicast beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 61, no. 10, pp. 2660–2674, May 2013.
- [12] O. Tervo, L. Tran, H. Pennanen, S. Chatzinotas, M. Juntti, and B. Ottersten, "Energy-efficient multi-cell multigroup multicasting with joint beamforming and antenna selection," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4904–4919, 2018.
- [13] H. A. Le Thi, V. N. Huynh, and T. P. Dinh, "DC programming and DCA for general DC programs," in *Advanced Computational Methods for Knowledge Engineering*. Cham: Springer International Publishing, 2014, pp. 15–35.
- [14] M. Hamid and I. Kostanic, "Path loss models for LTE and LTE-A relay stations," *Universal J. Commun. Netw.*, vol. 1, no. 4, pp. 119–126, 2013.