# Cost-Effectiveness Analysis and Design of Cost-Efficient Cell-Free Massive MIMO Systems

Wei Jiang* and Hans D. Schotten†
*German Research Center for Artificial Intelligence (DFKI)
Trippstadter Street 122, Kaiserslautern, 67663 Germany
†University of Kaiserslautern (RPTU)
Building 11, Paul-Ehrlich Street, Kaiserslautern, 67663 Germany

*Abstract*—Cell-free massive multi-input multi-output (MIMO) has recently attracted much attention, attributed to its potential to deliver uniform service quality. However, the adoption of a cell-free architecture raises concerns about the high implementation costs associated with deploying numerous distributed access points (APs) and the need for fronthaul network installation. To ensure the sustainability of next-generation wireless networks, it is crucial to improve cost-effectiveness, alongside achieving high performance. To address this, we conduct a cost analysis of cell-free massive MIMO and build a unified model with varying numbers of antennas per AP. Our objective is to explore whether employing multi-antenna APs could reduce system costs while maintaining performance. The analysis and evaluation result in the identification of a cost-effective design for cell-free massive MIMO, providing valuable insights for practical implementation.

## I. INTRODUCTION

In a cellular network, a base station (BS) is positioned at the center of a cell within a network of cells. High quality of service (QoS) is delivered to users at the cell center, close to the BS. However, the users at the cell edge experience worse QoS due to considerable distance-dependent path loss, strong inter-cell interference, and inherent handover issues within the cellular architecture. The performance gap between the cell center and edge is not merely a minor concern; it is substantial [1]. Recently, cell-free massive multi-input multi-output (MIMO) [2] has garnered much attention in both academia and industry due to its high potential for the upcoming sixth-generation (6G) systems [3]. There are no cells or cell boundaries. Instead, a multitude of distributed low-power, low-cost access points (APs) simultaneously serve users over the same time-frequency resource [4]. It perfectly matches some 6G scenarios, such as private or campus networks, with dedicated coverage areas like factories, stadiums, shopping malls, airports, railway stations, and exhibition halls. The cell-free architecture ensures uniform QoS for all users, effectively addressing the issue of under-served areas commonly encountered at the edges of conventional cellular networks [5].

Despite its considerable potential, the adoption of a cell-free architecture poses challenges related to high implementation costs associated with deploying numerous distributed APs and the necessity of installing a large-scale fronthaul network [6]. Deploying a traditional wireless network is already time-consuming and costly mainly due to acquiring and maintaining base station sites. In the cell-free system, this challenge intensifies as hundreds of suitable sites must be identified within a small area for wireless AP installations. Moreover, the installation of a large-scale fiber-cable network to interconnect these APs raises the expenditure and energy consumption [7]. To ensure the sustainability of next-generation wireless networks, improving the *cost-effectiveness* of cell-free systems is crucial, alongside achieving high performance.

To tackle this issue, we perform a cost analysis on cell-free massive MIMO, acknowledging that the overall cost including capital expenditure (CaPEX) and operational expenditure (OPEX) depends on the number of distributed wireless sites. Notably, the utilization of additional antennas at each AP does not induce extra site acquisition, and fiber connections, or incur supplementary maintenance costs. Consequently, reducing the number of APs by mounting multiple antennas on each AP can reduce implementation costs. However, a reduction in AP density compromises per-user spectral efficiency and sum capacity. This prompts us to seek a balance between performance and cost by determining the optimal number of antennas per AP. To provide a quantitative assessment for designing a cost-efficient architecture, a unified model encompassing cell-free architectures with varying antenna numbers per AP is built. We analyze and evaluate the spectral efficiency and cost-effectiveness in both downlink and uplink scenarios with maximal-ratio or zero-forcing schemes.

The structure of this paper is as follows: In Section II, the system and channel models are introduced. Sections III and IV focus on the analysis of uplink and downlink transmission, respectively. Section V presents the numerical results, and, lastly, Section VI draws the conclusions.

## II. CELL-FREE MASSIVE MIMO SYSTEM

In this paper, we argue that the cost-effectiveness of a cell-free massive MIMO system is proportionally related to the number of distributed AP sites. This is due to the fact that acquiring and maintaining AP sites incurs significant costs, and the installation of a large-scale fiber-cable network to

interconnect these APs amplifies the expenses. However, the fees associated with adding extra antennas to each AP are marginal since a unique advantage for massive MIMO systems is the use of low-cost antennas and RF components [8].

*1) System Model:* To offer a quantitative analysis for designing a cost-effective architecture, we formulate a general model for cell-free massive MIMO, allowing for the flexibility to set the number of antennas per AP. A total number of $M$ service antennas are distributed across $N_{AP}$ sites, where $1 < N_{AP} \leqslant M$. Each AP is equipped with $N_t$ antennas, adhering to $N_t \times N_{AP} = M$. These APs serve a few single-antenna user equipments (UEs) within a designated coverage area and the number of users $K \ll M$. In contrast to previous studies, our model accommodates a *variable* number of antennas per AP, covering conventional cell-free massive MIMO scenarios with $M$ distributed single-antenna APs ($N_t = 1$), like [2], [4], [9] and multi-antenna APs ($N_t > 1$) [10]. As depicted in Fig.1, a central processing unit (CPU) coordinates all APs through a fronthaul network to simultaneously serve all users over the same time-frequency resource.
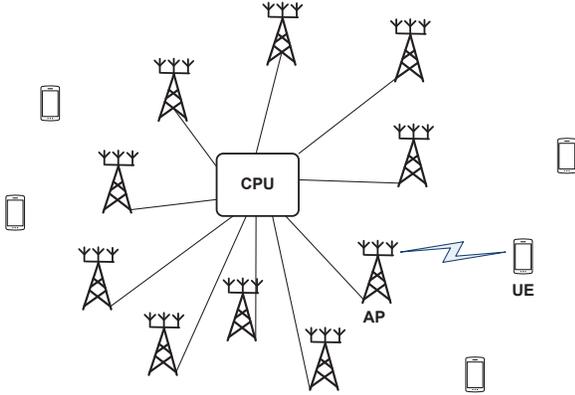


Fig. 1. A unified model for cell-free massive MIMO with a varying number of antennas per AP, where a total of $M$ antennas are distributed over $N_{AP}$ sites. If $N_{AP} = M$, it stands for typical cell-free massive MIMO with single-antenna APs, whereas we are interested in applying multi-antenna APs to improve cost-effectiveness by reducing the number of AP sites.

*2) Channel Model:* The channel coefficient connecting antenna $m$ (for all $m = 1, \ldots, M$) to UE $k$ (for all $k = 1, \ldots, K$) is represented as a circularly symmetric complex Gaussian random variable, denoted as $g_{mk} \in \mathcal{CN}(0, \beta_{mk})$. Here, $\beta_{mk}$ denotes large-scale fading, including path loss and shadowing. To alleviate the considerable downlink pilot overhead scaling with the number of service antennas, time-division duplex (TDD) is employed to separate the downlink and uplink signal transmission with the assumption of perfect channel reciprocity. Hence, each coherent interval is divided into three phases: uplink training, uplink data transmission, and downlink data transmission [4]. During the uplink training phase, UEs transmit orthogonal pilot sequences to acquire instantaneous channel state information (CSI). Using minimum mean-square error (MMSE) estimation [2], the network obtains channel estimates $\hat{g}_{mk}$, following a complex

normal distribution $\mathcal{CN}(0, \alpha_{mk})$ with $\alpha_{mk} = \frac{p_u \beta_{mk}^2}{p_u \beta_{mk} + \sigma_n^2}$, where $p_u$ represents the power of the UE transmitter, and $\sigma_n^2$ is the variance of noise [11]. The estimation error is defined as $\tilde{g}_{mk} = g_{mk} - \hat{g}_{mk}$, following the distribution of $\mathcal{CN}(0, \beta_{mk} - \alpha_{mk})$. In the case of multi-antenna APs, the large-scale fading between user $k$ and any antenna of the same AP $q$ (for $q = 1, \ldots, N_{AP}$) is identical. This assumption leads to $\beta_{mk} = \beta_{qk}$ and $\alpha_{mk} = \alpha_{qk}$, where $m \in \{(q-1)N_t+1, (q-1)N_t+2, \ldots, qN_t\}$.

## III. SPECTRAL-EFFICIENCY AND COST-EFFECTIVENESS ANALYSIS IN UPLINK TRANSMISSION

During uplink data transmission, each of the UEs aims to deliver its unit-variance, independent information-bearing symbol $x_k$ simultaneously to the APs, with UE $k$ transmitting $\sqrt{\eta_k}x_k$. The covariance matrix of the transmit vector $\mathbf{x} = [x_1, \ldots, x_K]^T$ adheres to $\mathbb{E}[\mathbf{xx}^H] = \mathbf{I}_K$, where $\mathbf{I}_K$ is the identity matrix of size $K$. The power coefficient is constrained by $0 \leqslant \eta_k \leqslant 1$. The received signals at AP $q$ can be expressed as an $N_t \times 1$ vector:

$$\mathbf{y}_q = \sqrt{p_u} \sum_{k=1}^{K} \mathbf{g}_{qk} \sqrt{\eta_k} x_k + \mathbf{n}_q, \qquad (1)$$

where $\mathbf{g}_{qk}$ denotes an $N_t \times 1$ channel signature between user $k$ and AP $q$ and the receiver noise $\mathbf{n}_q = [n_1, \ldots, n_{N_t}]^T$ follows a complex normal distribution $\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_{N_t})$.

In contrast to maximum likelihood, linear detection stands out for its lower complexity while still achieving commendable performance. Maximum-ratio combining (MRC), also known as matched filtering, is commonly employed for uplink detection in massive MIMO [12]. The underlying principle is to maximize the strength of the desired signal while disregarding inter-user interference (IUI) [13]. To detect the symbols, AP $q$ multiplies the received signals with the conjugate of its locally obtained channel estimates represented by $\hat{\mathbf{g}}_{qk}$, consisting of $\hat{g}_{mk}, \forall m \in \{(q-1)N_t+1, (q-1)N_t+2, \ldots, qN_t\}$. Then, the result $\hat{\mathbf{g}}_{qk}^* \otimes \mathbf{y}_q$, where $\otimes$ marks the Hadamard (element-wise) product, is delivered to the CPU via the fronthaual network.

As a result, the CPU observes

$$\mathbf{y} = \hat{\mathbf{G}}^H \left(\sqrt{p_u}\mathbf{GD}_\eta\mathbf{x} + \mathbf{n}\right)$$
$$= \hat{\mathbf{G}}^H \left(\sqrt{p_u} \sum_{k=1}^{K} \mathbf{g}_k \sqrt{\eta_k} x_k + \mathbf{n}\right). \qquad (2)$$

Here, $\mathbf{G}$ stands for an $M \times K$ channel matrix with the $(m, k)^{th}$ entry of $g_{mk}$, namely $[\mathbf{G}]_{mk} = g_{mk}$, $\hat{\mathbf{G}}$ is the matrix of channel estimates, i.e., $[\hat{\mathbf{G}}]_{mk} = \hat{g}_{mk}$, $\mathbf{g}_k$ represents the channel signature for user $k$ or the $k^{th}$ column of $\mathbf{G}$, and $\hat{\mathbf{g}}_k$ denotes the $k^{th}$ column of $\hat{\mathbf{G}}$. Additionally, $\mathbf{D}_\eta$ is a diagonal matrix given by $\mathbf{D}_\eta = \text{diag}([\eta_1, \ldots, \eta_K])$, and the receiver noise $\mathbf{n} = [n_1, \ldots, n_M]^T$ follows a complex normal distribution

$\mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_M)$. Decomposing (2) yields the received signal for detecting $x_k$ as

$$
\begin{aligned}
y_k &= \hat{\mathbf{g}}_k^H \left( \sqrt{p_u} \sum_{k=1}^K \mathbf{g}_k \sqrt{\eta_k} x_k + \mathbf{n} \right) \\
&= \sqrt{p_u \eta_k} \hat{\mathbf{g}}_k^H \mathbf{g}_k x_k + \sqrt{p_u} \sum_{i=1, i \neq k}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i \sqrt{\eta_i} x_i + \hat{\mathbf{g}}_k^H \mathbf{n}. \quad (3)
\end{aligned}
$$

Considering the vector of channel estimate errors denoted as $\tilde{\mathbf{g}}_k = \mathbf{g}_k - \hat{\mathbf{g}}_k$, (3) is further derived as

$$
\begin{aligned}
y_k &= \sqrt{p_u \eta_k} \hat{\mathbf{g}}_k^H (\hat{\mathbf{g}}_k + \tilde{\mathbf{g}}_k) x_k + \sqrt{p_u} \sum_{i=1, i \neq k}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i \sqrt{\eta_i} x_i + \hat{\mathbf{g}}_k^H \mathbf{n} \\
&= \sqrt{p_u \eta_k} \|\hat{\mathbf{g}}_k\|^2 x_k + \sqrt{p_u \eta_k} \hat{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k x_k \\
&\quad + \sqrt{p_u} \sum_{i=1, i \neq k}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i \sqrt{\eta_i} x_i + \hat{\mathbf{g}}_k^H \mathbf{n}. \quad (4)
\end{aligned}
$$

The CPU gets the full CSI exclusively when the CSI obtained from all APs is transmitted via the fronthaul network, or if the CPU conducts centralized estimation using observations also provided by the APs. Given the high signaling overhead, it is sensible to assume that the CPU is only equipped with knowledge of the channel statistics [2], [4]. That is to say, the CPU detects the received signals based on $\alpha_{mk}$, rather than $\hat{g}_{mk}, \forall m, k$. Thus, (4) is rewritten as

$$
\begin{aligned}
y_k &= \underbrace{\sqrt{p_u \eta_k} \mathbb{E}\left[\|\hat{\mathbf{g}}_k\|^2\right] x_k}_{\mathcal{S}_0: \, desired \, signal} + \underbrace{\sqrt{p_u \eta_k} \left(\|\hat{\mathbf{g}}_k\|^2 - \mathbb{E}\left[\|\hat{\mathbf{g}}_k\|^2\right]\right) x_k}_{\mathcal{I}_4: \, channel \, uncertainty \, error} \\
&+ \underbrace{\sqrt{p_u \eta_k} \hat{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k x_k}_{\mathcal{I}_1: \, channel \, estimation \, error} + \underbrace{\sqrt{p_u} \sum_{i=1, i \neq k}^K \hat{\mathbf{g}}_k^H \mathbf{g}_i \sqrt{\eta_i} x_i}_{\mathcal{I}_2: \, inter-user \, interference} \\
&+ \underbrace{\hat{\mathbf{g}}_k^H \mathbf{n}}_{\mathcal{I}_3: \, noise}. \quad (5)
\end{aligned}
$$

*a) Spectral-Efficiency Analysis:* The terms $\mathcal{S}_0$, $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, and $\mathcal{I}_4$ as defined in (5) exhibit mutual uncorrelation. As stated in [14], the worst-case noise for mutual information corresponds to Gaussian additive noise with a variance equal to the sum of the variances of $\mathcal{I}_1$, $\mathcal{I}_2$, $\mathcal{I}_3$, and $\mathcal{I}_4$. Thus, the uplink achievable rate for user $k$ is lower bounded by $R_k = \log(1 + \gamma_k^{ul})$, where

$$
\begin{aligned}
\gamma_k^{ul} &= \frac{\mathbb{E}\left[|\mathcal{S}_0|^2\right]}{\mathbb{E}\left[|\mathcal{I}_1 + \mathcal{I}_2 + \mathcal{I}_3 + \mathcal{I}_4|^2\right]} \\
&= \frac{\mathbb{E}\left[|\mathcal{S}_0|^2\right]}{\mathbb{E}\left[|\mathcal{I}_1|^2\right] + \mathbb{E}\left[|\mathcal{I}_2|^2\right] + \mathbb{E}\left[|\mathcal{I}_3|^2\right] + \mathbb{E}\left[|\mathcal{I}_4|^2\right]} \quad (6)
\end{aligned}
$$

with

$$
\mathbb{E}\left[|\mathcal{S}_0|^2\right] = p_u \eta_k N_t^2 \left( \sum_{q=1}^{N_{AP}} \alpha_{qk} \right)^2 \quad (7)
$$

$$
\mathbb{E}\left[|\mathcal{I}_1|^2\right] = p_u \eta_k N_t \sum_{q=1}^{N_{AP}} (\beta_{qk} - \alpha_{qk}) \alpha_{qk} \quad (8)
$$

$$
\mathbb{E}\left[|\mathcal{I}_2|^2\right] = p_u N_t \sum_{i=1, i \neq k}^K \eta_i \sum_{q=1}^{N_{AP}} \beta_{qi} \alpha_{qk} \quad (9)
$$

$$
\mathbb{E}\left[|\mathcal{I}_3|^2\right] = \sigma_n^2 N_t \sum_{q=1}^{N_{AP}} \alpha_{qk} \quad (10)
$$

$$
\mathbb{E}\left[|\mathcal{I}_4|^2\right] = N_t \sum_{q=1}^{N_{AP}} \alpha_{qk}^2. \quad (11)
$$

Substituting the above terms into (6), yields

$$
\gamma_k^{ul} = \frac{p_u \eta_k N_t^2 \left(\sum_{q=1}^{N_{AP}} \alpha_{qk}\right)^2}{p_u N_t \sum_{i=1}^K \eta_i \sum_{q=1}^{N_{AP}} \alpha_{qk} \beta_{qi} + \sigma_n^2 N_t \sum_{q=1}^{N_{AP}} \alpha_{qk}}. \quad (12)
$$

By now, we get the closed-form expression of uplink per-user spectral efficiency for a generalized system with $1 < N_{AP} \leqslant M$ single- or multi-antenna APs.

*b) Cost Analysis:* In evaluating the cost-effectiveness of the system, the overall cost of a cell-free massive MIMO system is formulated as

$$
\begin{aligned}
C &= c_{ls} + c_{cpu} + c_{mo} \\
&+ N_{AP}\Big[ c_{sc} + c_{ps} + c_{fb} + c_{bb} + N_t(c_{ant} + c_{rf}) \Big]. \quad (13)
\end{aligned}
$$

This comprehensive cost includes various components such as $c_{ls}$ representing fees associated with acquiring spectrum licenses from regulatory authorities, $c_{cpu}$ denoting the expense of constructing the CPU, and $c_{mo}$ covering the costs of electricity consumption, routine maintenance, network management, monitoring, and insurance. Furthermore, $c_{sc}$ accounts for the expenses related to obtaining or leasing space for a wireless site and site construction, $c_{fb}$ addresses the establishment cost of a fiber optic connection to the CPU, $c_{ps}$ encompasses the installation cost of power supply infrastructure, $c_{bb}$ represents the cost of a baseband unit, while $c_{ant}$ and $c_{rf}$ correspond to the cost per antenna and the cost of the RF chain, respectively.

To investigate the influence of the number of antennas per AP on cost-effectiveness, we redefine (13) in a new form as:

$$
C = N_{AP}\Big[ C_f + N_t C_v \Big]. \quad (14)
$$

Here, the combined costs of $c_{ls} + c_{cpu} + c_{mo}$ are equally allocated to each AP by dividing by $N_{AP}$, $C_f$ denotes the fixed cost independent of the number of AP antennas

$$
C_f = \frac{c_{ls} + c_{cpu} + c_{mo}}{N_{AP}} + c_{sc} + c_{ps} + c_{fb} + c_{bb}, \quad (15)
$$

and the fees associated with the number of AP antennas are represented by $C_v = c_{ant} + c_{rf}$. Consequently, the cost-effectiveness of the system in the uplink can be assessed through the ratio between the sum rate and the overall cost, i.e.,

$$\Gamma_{ul} = \frac{\sum_{k=1}^{K} \log(1 + \gamma_k^{ul})}{N_{AP}(C_f + N_t C_v)}. \tag{16}$$

## IV. SPECTRAL-EFFICIENCY AND COST-EFFECTIVENESS ANALYSIS IN DOWNLINK TRANSMISSION

In contrast to theoretically optimal but impractical approaches like dirty-paper coding, linear precoding stands out for its appealing combination of low complexity and effective performance. In cell-free massive MIMO systems, the spatial multiplexing of information symbols is commonly realized using two linear precoding techniques: conjugate beamforming (CBF) [2] and zero-forcing precoding (ZFP) [15]. The symbols intended for $K$ users is denoted by $\mathbf{u} = [u_1, \ldots, u_K]^T$, where $\mathbb{E}[\mathbf{u}\mathbf{u}^H] = \mathbf{I}_K$. Let $\mathbf{B}$ represent the $M \times K$ precoding matrix with elements $[\mathbf{B}]_{mk} = \sqrt{\eta_{mk}} b_{mk}$, where $\eta_{mk}$ denotes the power coefficient for the $k^{th}$ user at antenna $m$, and $b_{mk}$ is the precoding coefficient. Adhering to the per-antenna power constraint $p_d$ and a noise vector $\mathbf{w} = [w_1, \ldots, w_K]^T \in \mathcal{CN}(\mathbf{0}, \sigma_n^2 \mathbf{I}_K)$, the collective representation of received symbols for all users, namely $\mathbf{r} = [r_1, \ldots, r_K]^T$, is given by:

$$\mathbf{r} = \sqrt{p_d} \mathbf{G}^T \mathbf{B} \mathbf{u} + \mathbf{w}. \tag{17}$$

Equivalently, the $k^{th}$ user has the observation of

$$
\begin{aligned}
r_k &= \sqrt{p_d} \mathbf{g}_k^T \mathbf{B} \mathbf{u} + w_k \\
&= \sqrt{p_d} \mathbf{g}_k^T \sum_{i=1}^{K} \mathbf{b}_i u_i + w_k \\
&= \sqrt{p_d} \mathbf{g}_k^T \mathbf{b}_k u_k + \sqrt{p_d} \sum_{i=1, i \neq k}^{K} \mathbf{g}_k^T \mathbf{b}_i u_i + w_k,
\end{aligned} \tag{18}
$$

where $\mathbf{b}_k \in \mathbb{C}^{M \times 1}$ is the $k^{th}$ column of $\mathbf{B}$.

*1) Conjugate Beamforming:* Similar to the MRC technique employed in the uplink, CBF is designed to optimize the reception of the desired signal [2], aiming to maximize its strength. Its precoding matrix is given by $[\mathbf{B}]_{mk} = \sqrt{\eta_{mk}} \hat{g}_{mk}^*$. Applying CBF, (18) can be rewritten into an element-wise form as

$$
\begin{aligned}
r_k &= \sqrt{p_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} g_{mk} \hat{g}_{mk}^* u_k \\
&+ \sqrt{p_d} \sum_{i=1, i \neq k}^{K} \sum_{m=1}^{M} \sqrt{\eta_{mi}} g_{mk} \hat{g}_{mi}^* u_i + w_k.
\end{aligned} \tag{19}
$$

Each user only knows channel statistics $\mathbb{E}\left[|\hat{g}_{mk}|^2\right] = \alpha_{mk}$ rather than channel estimate $\hat{g}_{mk}$ since there are no downlink pilots. As a result, each user detects the signals based on channel statistics. Like (5), (19) is transformed to

$$
\begin{aligned}
r_k &= \sqrt{p_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \mathbb{E}[|\hat{g}_{mk}|^2] u_k \\
&+ \sqrt{p_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \left(|\hat{g}_{mk}|^2 - \mathbb{E}[|\hat{g}_{mk}|^2]\right) u_k \\
&+ \sqrt{p_d} \sum_{m=1}^{M} \sqrt{\eta_{mk}} \tilde{g}_{mk} \hat{g}_{mk}^* u_k \\
&+ \sqrt{p_d} \sum_{i=1, i \neq k}^{K} \sum_{m=1}^{M} \sqrt{\eta_{mi}} g_{mk} \hat{g}_{mi}^* u_i + w_k.
\end{aligned} \tag{20}
$$

The power control is generally decided by site-specific large-scale fading coefficient $\beta_{qk}$, we have $\eta_{mk} = \eta_{qk}$ for all $m \in \{(q-1)N_t + 1, (q-1)N_t + 2, \ldots, qN_t\}$. Applying analogous manipulations to the derivation of (12), we obtain the effective SINR as

$$\gamma_k^{cbf} = \frac{p_d N_t^2 \left(\sum_{q=1}^{N_{AP}} \sqrt{\eta_{qk}} \alpha_{qk}\right)^2}{\sigma_n^2 + p_d N_t \sum_{q=1}^{N_{AP}} \beta_{qk} \sum_{i=1}^{K} \eta_{qi} \alpha_{qi}}. \tag{21}$$

Let $N_{AP} = M$ and $N_t = 1$, (21) reverts to equation (27) in [2], illustrating the performance of conventional cell-free massive MIMO with single-antenna APs. Nevertheless, we progress beyond by covering the scenario of multi-antenna APs when $N_t > 1$. Like (16), the cost-effectiveness of CBF-based cell-free massive MIMO in the downlink can be evaluated by

$$\Gamma_{dl}^{cbf} = \frac{\sum_{k=1}^{K} \log(1 + \gamma_k^{cbf})}{N_{AP}(C_f + N_t C_v)}. \tag{22}$$

*2) Zero-Forcing Precoding:* Instead of maximizing the strength of the desired signal, ZFP aims to cancel IUI at the UE receiver [16]. Its precoding matrix is the pseudo inverse of the channel matrix, i.e., $\mathbf{B} = \hat{\mathbf{G}}^* (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \otimes \mathbf{E}$, where $[\mathbf{E}]_{mk} = \sqrt{\eta_{mk}}$. As proved by [4], it is necessary to have $\eta_{1k} = \eta_{2k} = \cdots = \eta_{Mk}, \forall k$ to keep $\hat{\mathbf{G}}^T \mathbf{B}$ orthogonal such that the IUI is eliminated. Hence, we have $\eta_{mk} = \eta_k, \forall m$ and therefore $\mathbf{B} = \hat{\mathbf{G}}^* (\hat{\mathbf{G}}^T \hat{\mathbf{G}}^*)^{-1} \mathbf{D}$, where $\mathbf{D} = \text{diag}([\eta_1, \ldots, \eta_K])$. The IUI term in (18) is zero-forced, namely $\sum_{i=1, i \neq k}^{K} \mathbf{g}_k^T \mathbf{b}_i u_i = 0$, yielding $r_k = \sqrt{p_d} \mathbf{g}_k^T \mathbf{b}_k u_k + w_k$, which is further derived to

$$
\begin{aligned}
r_k &= \sqrt{p_d} \hat{\mathbf{g}}_k^T \mathbf{b}_k u_k + \sqrt{p_d} \tilde{\mathbf{g}}_k^T \mathbf{b}_k u_k + w_k \\
&= \sqrt{p_d \eta_k} u_k + \sqrt{p_d} \tilde{\mathbf{g}}_k^T \mathbf{b}_k u_k + w_k.
\end{aligned} \tag{23}
$$

The effective SINR for user $k$ can be expressed as follows:

$$\gamma_k^{zfp} = \frac{p_d \eta_k}{\sigma_n^2 + p_d \sum_{i=1}^{K} \eta_i \chi_i^k}. \tag{24}$$

where $\chi_i^k$ represents the $i^{th}$ diagonal element of the $K \times K$ matrix dedicated to user $k$ [15]:

$$\mathbb{E}\left[\left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1} \hat{\mathbf{G}} \mathbb{E}\left[\tilde{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k\right] \hat{\mathbf{G}}^H \left(\hat{\mathbf{G}}\hat{\mathbf{G}}^H\right)^{-1}\right], \tag{25}$$
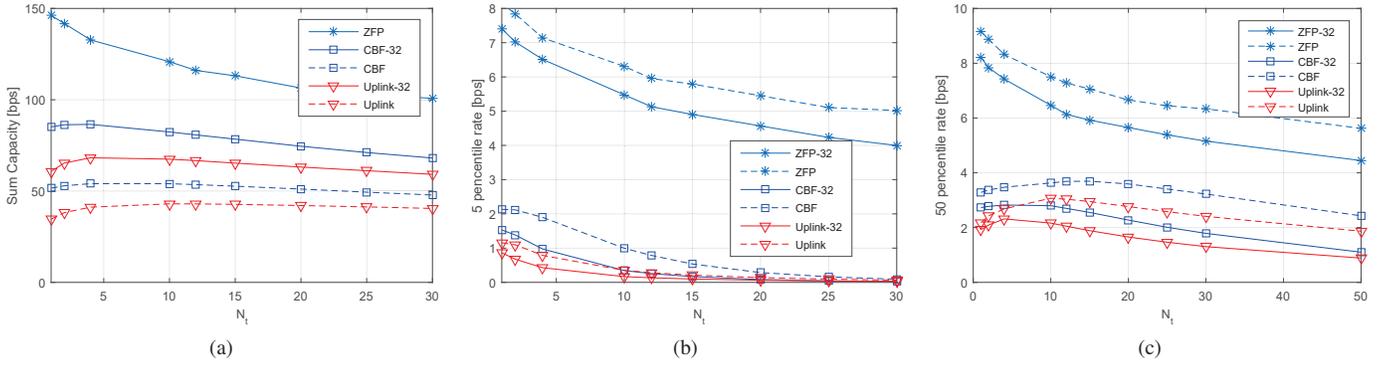
Fig. 2. Performance of cell-free massive MIMO systems in terms of the number of antennas at each AP (i.e., $N_t$), where a total of $M = 300$ antennas serve $K = 16$ (default) or $K = 32$ users (marked by -32 in the legend), including (a) the sum capacity; (b) the 5-percentile spectral efficiency; and (b) the 50-percentile spectral efficiency.

and $\mathbb{E}\left[\tilde{\mathbf{g}}_k^H \tilde{\mathbf{g}}_k\right]$ is a diagonal matrix with the $k^{th}$ diagonal element equaling to $\beta_{mk} - \alpha_{mk}$. The cost-effectiveness of ZFP-based cell-free massive MIMO can be assessed through

$$\Gamma_{dl}^{zfp} = \frac{\sum_{k=1}^{K} \log(1 + \gamma_k^{zfp})}{N_{AP}(C_f + N_t C_v)}. \qquad (26)$$

## V. NUMERICAL RESULTS

The spectral efficiency and cost-effectiveness of a cell-free massive MIMO system are numerically evaluated in terms of the number of antennas per AP. Consider a configuration where a total of $M = 300$ antennas serve users across a square area of $1 \times 1\text{km}^2$. The large-scale fading is figured out by the formula $\beta_{mk} = 10^{\frac{\mathcal{L}_{mk} + \mathcal{X}_{mk}}{10}}$. The shadowing $\mathcal{X}_{mk}$, commonly modeled as a log-normal distribution $\mathcal{N}(0, \sigma_{sd}^2)$, where $\sigma_{sd} = 8\text{dB}$. The path loss, determined by the COST-Hata model [2], is expressed as

$$\mathcal{L}_{mk} = \begin{cases} -L_0 - 35\log_{10}(d_{mk}), & d_{mk} > d_1 \\ -L_0 - 10\log_{10}(d_1^{1.5}d_{mk}^2), & d_0 < d_{mk} \leq d_1 \\ -L_0 - 10\log_{10}(d_1^{1.5}d_0^2), & d_{mk} \leq d_0 \end{cases}$$
$$(27)$$

where $d_{mk}$ is the distance between user $k$ and antenna $m$, the three-slope breakpoints are set as $d_0 = 10\text{m}$ and $d_1 = 50\text{m}$, and $L_0 = 140.72\text{dB}$ is defined by

$$L_0 = 46.3 + 33.9\log_{10}(f_c) - 13.82\log_{10}(h_{AP}) \qquad (28)$$
$$- [1.1\log_{10}(f_c) - 0.7]h_{UE} + 1.56\log_{10}(f_c) - 0.8$$

Here, the carrier frequency is $f_c = 1.9\text{GHz}$, the height of the AP antenna is $h_{AP} = 15\text{m}$, and the height of the UE is $h_{UE} = 1.65\text{m}$.

The per-antenna and UE power constraints are $p_d = p_u = 200\text{mW}$. The white noise power density is $-174\text{dBm/Hz}$ with a noise figure of 9dB, and the signal bandwidth is set at 5MHz. In the uplink, performed in a distributed manner, it is reasonable for each UE to adopt a full-power strategy with $\eta_k = 1$. Regrettably, the optimal max-min power-control schemes in the downlink, employing both

CBF and ZFP, are too computationally complex for practical implementation. In line with the recommendation by [4], we opt for sub-optimal schemes characterized by lower complexity. To elaborate, within the ZFP approach, we set $\eta_1 = \ldots = \eta_K = \left(\max_m \sum_{k=1}^{K} \delta_{km}\right)^{-1}$, where $\boldsymbol{\delta}_m = [\delta_{1m}, \ldots, \delta_{Km}]^T = \text{diag}(\mathbb{E}[(\hat{\mathbf{G}}\hat{\mathbf{G}}^H)^{-1}\hat{\mathbf{g}}_m\hat{\mathbf{g}}_m^H\hat{\mathbf{G}}\hat{\mathbf{G}}^H)^{-1}])$ and $\hat{\mathbf{g}}_m$ represents the $m^{th}$ column of $\hat{\mathbf{G}}$. In the case of CBF, the APs employ a full-power strategy, mathematically denoted as $\eta_m = (\sum_{k=1}^{K} \alpha_{mk})^{-1}, \forall m$.

During the simulations, the number of AP antennas varies as $N_t \in \{1, 2, 4, 10, 12, 15, 20, 25, 30, 50\}$. Correspondingly, a total of 300 antennas were distributed to $N_{AP} \in \{300, 150, 75, 30, 25, 20, 15, 10, 6\}$ APs. Two groups of simulations were performed with the number of users set to $K = 16$ and $K = 32$, respectively. Fig.2a illustrates the sum rate for three scenarios: uplink, downlink with ZFP, and downlink with CBF. As expected, in the downlink, ZFP outperforms CBF significantly due to ZFP's utilization of global CSI to eliminate inter-user interference for all users. In contrast, CBF relies on only local CSI for precoding. The superiority of ZFP comes at the cost of high signaling overhead, as the CSI must be delivered via the fronthaul network. While both CBF in the downlink and MRC in the uplink share the approach of maximizing the desired signal, CBF's performance surpasses that of MRC. This is attributed to the higher power consumption in the downlink, amounting to $M \times p_t = 60\,\text{W}$, compared to the uplink's $K \times p_u = 3.2\,\text{W}$ or $6.4\,\text{W}$. To provide a clear illustration, we omit the ZFP curve with 32 users as it deviates significantly from the curves of CBF and MRC. As observed, the sum rate of ZFP monotonically decreases with the increasing number of AP antennas. However, in the case of CBF and MRC, optimal performance is achieved using multi-antenna APs at $N_t = 4$.

The concept of user-experienced data rate, as defined by 3GPP, is rooted in the $5^{th}$ percentile point (5%) of the cumulative distribution function of user throughput. This metric provides a meaningful measurement of perceived performance, particularly at the cell edge. Fig.2b presents the $5^{th}$ percentile per-user spectral efficiency, offering a glimpse into how the
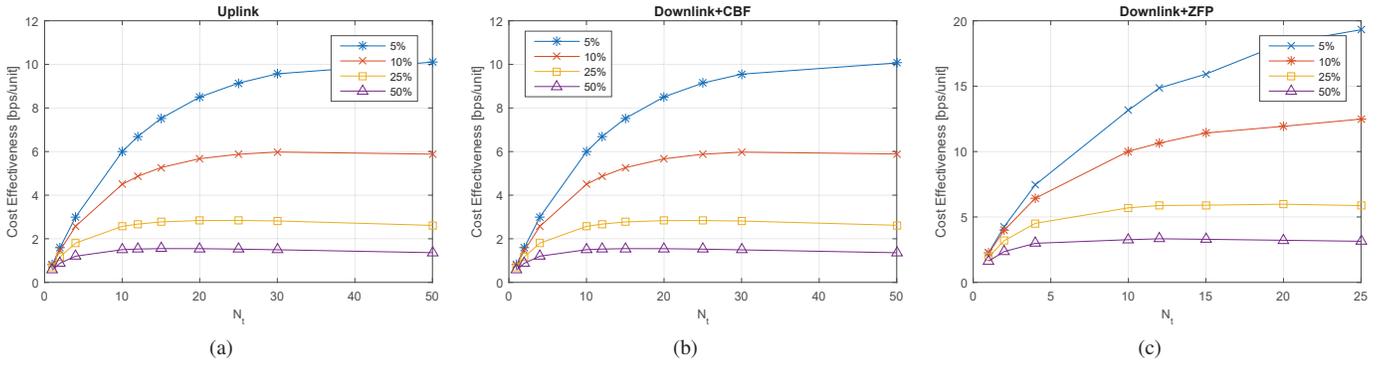
Fig. 3. Cost-effectiveness results of cell-free massive MIMO systems in terms of the number of antennas at each AP (i.e., $N_t$), where a total of $M = 300$ antennas serve $K = 16$. The ratio between the $N_t$-dependent and $N_t$-independent costs takes four values, i.e., $C_v/C_f = 0.05, 0.1, 0.25$ and $0.5$.

cell-edge performance varies with $N_t$. Simultaneously, Fig.2c displays the $50^{th}$ percentile, or median, per-user spectral efficiency. Observing Fig.2b, it becomes evident that the cell-edge performance is sensitive to reductions in AP density. However, a crucial insight emerges: the majority of performance levels can be maintained when the number of AP antennas remains moderate, specifically when $N_t < 5$. In contrast, the median spectral efficiency demonstrates greater adaptability to multi-antenna AP scenarios. The peak performance for CBF and MRC occurs when there are approximately $N_t = 10$ antennas. While the performance of ZFP exhibits a monotonically decreasing trend with the increasing number of AP antennas, it performs still well when $N_t$ is high.

We evaluated cost-effectiveness based on the achievable spectral efficiency of each AP per cost unit, varying the number of antennas at each AP, where a total of $M = 300$ antennas serve $K = 16$. To maintain generality, we set the costs independent of $N_t$ to a fixed value of one cost unit during all simulations ($C_f = 1$). Across specific deployment scenarios, the fee for adding an extra antenna and its associated RF chain varies from $5\%$ to $50\%$ of the fixed cost. In particular, the ratio between $N_t$-dependent and $N_t$-independent costs takes four values: $C_v/C_f = 0.05, 0.1, 0.25$, and $0.5$. The results, as depicted in the Fig.3, show a decrease as the cost of $C_v$ increases, reflecting an elevation in the total deployment cost without necessarily indicating a decline in performance. It is crucial to note that, regardless of downlink or uplink, and whether using ZFP or CBF, their cost-effectiveness monotonically grows with the increasing number of AP antennas.

## VI. CONCLUSIONS

This paper investigated the cost-effectiveness aspect of cell-free massive MIMO through comprehensive cost and performance analysis. Our unified model, accommodating varying numbers of antennas in both uplink and downlink scenarios, shed light on the impact of multi-antenna APs on economic viability. Factors like site acquisition, fiber connection, maintenance, and hardware costs were examined. The findings underscored the efficacy of multi-antenna APs in enhancing cost-effectiveness, albeit with a trade-off: reducing AP density

may compromise per-user spectral efficiency and sum capacity. However, we demonstrated that the majority of performance levels can be maintained by determining the suitable number of antennas per AP. The analysis and evaluation result in the identification of a cost-effective design for cell-free massive MIMO, providing valuable insights for practical implementation.

## REFERENCES

[1] W. Jiang and H. D. Schotten, "Cell-edge performance booster in 6G: Cell-free massive MIMO vs. reconfigurable intelligent surface," in *Proc. IEEE Eur. Conf. on Netw. and Commun. (EUCNC)*, Gothenburg, Sweden, Jun. 2023, pp. 1–6.

[2] H. Q. Ngo *et al.*, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.

[3] W. Jiang *et al.*, "The road towards 6G: A comprehensive survey," *IEEE Open J. Commun. Society*, vol. 2, pp. 334–366, Feb. 2021.

[4] E. Nayebi *et al.*, "Precoding and power optimization in cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 16, no. 7, pp. 4445–4459, Jul. 2017.

[5] S. Buzzi *et al.*, "User-centric 5G cellular networks: Resource allocation and comparison with the cell-free massive MIMO approach," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1250–1264, Feb. 2020.

[6] H. Masoumi and M. J. Emadi, "Performance analysis of cell-free massive MIMO system with limited fronthaul capacity and hardware impairments," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1038–1052, Feb. 2020.

[7] W. Jiang and H. D. Schotten, "Hierarchical cell-free massive MIMO for high capacity with simple implementation," in *Proc. 2024 IEEE Int. Commun. Conf. (ICC)*, Denver, USA, Jun. 2024, pp. 1–6.

[8] T. L. Marzetta, "Massive MIMO: An introduction," *Bell Labs Technical J.*, vol. 20, pp. 11 – 22, Mar. 2015.

[9] H. Q. Ngo *et al.*, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.

[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] W. Jiang and H. D. Schotten, "Cell-free massive MIMO-OFDM transmission over frequency-selective fading channels," *IEEE Commun. Lett.*, vol. 25, no. 8, pp. 2718 – 2722, Aug. 2021.

[12] W. Jiang and F.-L. Luo, "Cellular and cell-free massive MIMO techniques in 6G," in *6G Key Technologies: A Comprehensive Guide*. New York, USA: John Wiley&Sons and IEEE Press, 2023, ch. 9.

[13] W. Jiang and H. Schotten, "Opportunistic AP selection in cell-free massive MIMO-OFDM systems," in *Proc. 2022 IEEE 95th Veh. Techno. Conf. (VTC2022-Spring)*, Helsinki, Finland, Jun. 2022.

[14] 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.

[15] W. Jiang and H. Schotten, "Impact of channel aging on zero-forcing precoding in cell-free massive MIMO systems," *IEEE Commun. Lett.*, vol. 25, no. 9, pp. 3114 – 3118, Sep. 2021.

[16] ——, "Deep learning-aided delay-tolerant zero-forcing precoding in cell-free massive MIMO," in *Proc. 2022 IEEE 96th Veh. Techno. Conf. (VTC2022-Fall)*, London, UK, Sep. 2022.