Decentralized Distributed Massive MIMO

Junbo Zhao[1], Mostafa Rahmani[1], and Alister G Burr[1]

[1]School of Physics, Engineering and Technology, University of York

February 07, 2025

Decentralized Distributed Massive MIMO

Junbo Zhao, Mostafa Rahmani, and Alister G. Burr, *Senior Member, IEEE*

School of Physics, Engineering and Technology, University of York, UK

Abstract—In this paper, we introduce the concept of "Decentralized Distributed Massive MIMO" (multiple-input multiple-output), which decentralizes the baseband processing to the network edge and distributes the Massive MIMO (also known as cell-free massive MIMO). In Decentralized Distributed Massive MIMO (DD-mMIMO), the signal processing is performed at edge processing units (EPUs) located closer to the access points (APs) rather than in a single remote central processing unit (CPU), significantly reducing the latency. This avoids cluster edge effects that may arise in cell-free massive MIMO. To avoid this, DD-mMIMO defines coordination regions which may overlap, which implies the APs in the overlap area may be coordinated by more than one EPU. To analyse the performance of the system, we derive an expression of the achievable uplink spectral efficiency (SE) within multiple-antenna APs for various data detection algorithms. The numerical results demonstrate that DD-mMIMO has the potential to fulfill the ultra-dense low-latency requirement in the next generation mobile network and beyond.

Index Terms—AP coordination, beyond mobile network, cell-free massive MIMO, Decentralized Distributed Massive MIMO, ultra-dense network.

I. Introduction

MASSIVE multiple-input multiple-output (mMIMO) originally proposed in [1] has been recognized as an important technique in the fifth generation (5G) wireless communications [2]–[4], due to the need for a greatly increased data rate. The key feature of mMIMO is that each base station, which is equipped with many more antennas than there are user terminals (UTs), simultaneously serves all users [2]. This improves the spectral efficiency (SE) while requiring only updates to the hardware of the present base stations rather than new base station deployments [5]. There are two schemes to deploy the massive antennas for the base stations. One is the collocated mMIMO system where the multiple antennas are embodied in one base station, which has the advantage of eliminating the multipath fading. That is because as the number of antennas at the base station increases, channel hardening appears which means the channel becomes more deterministic than random [6]. However, in collocated mMIMO, coverage may be limited leading to poor data rates at the edge of the cell, (see fig. 3 in [7]). Another scheme is distributed mMIMO [8]–[10], where many antennas are distributed over multiple access points (APs) equipped with single or a few antennas which are closer to UTs, so the probability of coverage increases. The approach in which many APs simultaneously serve UTs across the whole network area without any cell or cell boundaries, is called cell-free mMIMO [2]. Since the concept of cell vanishes, there are no disadvantaged cell edge users.

Signal processing techniques, such as maximum ratio combining (MRC) (a.k.a matched filter), zero-forcing (ZF) and minimum mean square error (MMSE) processing [11], [12], can be easily used in the mMIMO system. The paper [2] investigated local MRC at each AP, [13] evaluated fully centralized MRC at the CPU, while [5] analyzed the local and fully centralized MMSE method. Among these combining techniques, MMSE processing gives the best SE performance in the cell-free mMIMO network however it has a high computation complexity. To fulfill the demands of MMSE processing, cell-free mMIMO can be deployed using cloud-RAN (C-RAN).

A. Radio Access Network (RAN) Architecture

RAN architectures such as C-RAN, Fog-RAN (F-RAN) and virtualized RAN (vRAN) have been successively proposed in the last decade. C-RAN, which may apply a coordinated multipoint (CoMP) approach to address the intercellular interference, is a promising network architecture in 5G [14], [15]. It is composed of the remote radio head (RRH) and the baseband processing unit (BBU) at which the quantized signal is processed, connected by high bandwidth "fronthaul" connections [15], [16]. The BBUs are then collected in a "cloud" (which is equivalent to the CPU in cell-free mMIMO) providing the high computation required. However, in C-RAN, the large fronthaul load and the highly centralized computation limit network scalability and may cause significant delay, so that the requirements for 5G ultra-reliable low-latency communications (URLLC) [17] cannot be achieved. Thus, a new paradigm known as F-RAN has been proposed, with its core concept involving the shift of signal processing from the "cloud" back to the APs at the network edge [14]. F-RAN has been discussed in [18], [19] which shows that using an edge processing unit (EPU) in place of the remote CPU can improve the latency. In the past few years, the Open RAN concept has been presented by the O-RAN Alliance [20]. It aims to standardize the architecture and interfaces of the RAN [21] to open the RAN to new providers and improve its flexibility. The Open RAN architecture disaggregates the base stations between three functional units: Central Unit (CU), Distributed Unit (DU) and Radio Unit (RU) [21], and the functions of the physical layer are divided between the RU and DU according to options defined by the Third Generation Partnership Project (3GPP) [22]. The open interfaces between these components enable the deployment of the RAN architecture with different network location selections (cloud, edge, cellular base station) [21], [23].

B. The Scalable Network

The conventional cell-free mMIMO system described in [2], where all APs are connected to a single CPU to serve all UTs simultaneously, is inherently not scalable. This scalability issue arises because, as the network coverage expands indefinitely, the number of fronthaul links and the fronthaul distances also grow unbounded, rendering such an implementation impractical. Furthermore, if a user is served by all APs, then the computation per user will also tend to infinity. Scalability has been discussed in [24], [25], where it is defined as ensuring that the computational complexity per AP remains finite as the number of UTs in the entire network grows infinitely large. Furthermore, mitigating pilot contamination caused by non-orthogonal pilots requires knowledge of large-scale fading from all APs to be available at the CPU [26]. However, this approach is not scalable, particularly in ultra-dense low-latency networks with dynamic mobility. Similarly, [27] highlights the scalability challenge, noting that increased coordination levels among APs lead to higher computational complexity, latency, and fronthaul capacity requirements.

Several approaches have been proposed to achieve scalable networks. One such method, presented in [24], involves the formation of dynamic cooperation clusters, where a limited number of APs serve a specific user within the cluster, under the assumption that each AP serves at most one user per pilot sequence. Additionally, in [24] the partial MMSE (P-MMSE) and local P-MMSE (LP-MMSE) combining techniques are utilized for data detection, ensuring the scalability of the network. However, these combining techniques come with a trade-off, as they partially reduce the ability to mitigate interference from other users. Moreover, this kind of scalability is addressed using a user-centric framework with dynamic cooperation clustering (DCC), which may not be practical and flexible, as the cluster can change rapidly with user access and mobility [28]. An alternative, simpler approach adopts the division of the cell-free mMIMO network into separate regions, each managed by independent CPUs, as discussed in [25]. However,the edge effect arises in this method, leading to interference between adjacent clusters, as highlighted in [14]. To mitigate the interference, cooperation techniques for the edge APs should be considered. To overcome these challenges, this paper proposes a novel scalable network architecture.

C. Motivations

In this paper, our fundamental idea, beyond the concepts in [24], is to explicitly define decentralized locations nearer to the network edge where the signal processing for specific UTs will take place. Thus processing is decentralized, and does not occur at the cloud or center. To solve the problem of inter-cluster interference mentioned above, we define adjacent entities, referred to as coordination regions, which naturally overlap, and that all UTs within these regions are coordinated by the corresponding EPU. This allows APs near the edge of a coordinating region to be connected to multiple EPUs, thus avoiding the edge effect. Based on the features of the technologies discussed above, we integrate decentralized signal processing with cell-free mMIMO, defining coordination

regions and proposing a new paradigm called "Decentralized Distributed Massive MIMO". This system is derived from the "Fog Massive MIMO" concept introduced in [14]. The name change reflects shifts in terminology over the past six years. Originally, "Fog Massive MIMO" was not widely used; however, in recent years, the term has appeared in [26] to describe networks with high density of RRHs (which is called the "fog") and the UTs autonomously connect to the most convenient RRH: this is different from [14]. Additionally, the term "fog" is commonly associated with fog/edge computing, which involves network caching. Therefore, we avoid the term "fog" in this context and instead emphasize decentralization of signal processing, and combine this with the distributed nature of cell-free mMIMO. A simple performance analysis for the uplink DD-mMIMO has been given in [14], using maximum ratio combining (MRC) for the data detection with least squares (LS) channel estimation and adopting the assumption of channel hardening. However this might not be valid, since the number of antennas at APs is small or the path-loss exponent is large (see [29] for a discussion of the applicability of channel hardening in cell-free mMIMO). So, in this paper, we discard this assumption and primarily consider the general case where each AP is equipped with either a single antenna or a small number of antennas (up to eight), and using both MRC and MMSE receive combining techniques. To provide a comprehensive analysis and fair comparison of cell-free mMIMO and DD-mMIMO, we consider an ultra-dense network where the density of APs is four times that of UTs for both systems and use the propagation model specified by the 3GPP standard. Power control is not considered in this paper; instead, equal power is allocated to all transmitting UTs.

D. Contributions

The main contributions of this paper are summarized as follows:

- We propose a new architecture "Decentralized Distributed Massive MIMO" which emphasizes that the signal processing is carried out in the EPUs at the network edge. Unlike the canonical cell-free mMIMO system, where all APs are connected to one CPU, DD-mMIMO separates the whole network into multiple clusters, each managed by an EPU. This design alleviates the computational burden on a single CPU, decreases latency, and reduces optical fiber costs due to the shorter required transmission distance on the fronthaul links.
- We define the coordination region for the DD-mMIMO system as the area in which APs are coordinated by one or more EPUs. This setup leverages enhanced information sharing among coordinated APs, thereby improving performance for UTs at the cluster edge.
- We present a practical pilot allocation method for the DD-mMIMO system, ensuring that pilot sequences assigned within each coordination region are fully orthogonal. This method leverages coordination among UTs to significantly mitigate interference from pilot contamination.
- We discuss the scalability of the DD-mMIMO system and compare it to the scalable cell-free mMIMO definition

provided in [24], demonstrating that our proposed architecture serves as an alternative scalable cell-free mMIMO model. Unlike the DCC approach in [24], [25] where serving APs are dynamically selected by each UT, DD-mMIMO allows greater flexibility in system deployment, even with rapid location changes of UTs.

- We derive rigorous expressions for the instantaneous uplink signal-to-interference-and-noise ratio (SINR) and achievable SE in the DD-mMIMO system. These expressions are valid for APs with an arbitrary number of antennas, account for imperfect channel estimation, and support various receive combining techniques.
- We compare the uplink performance of cell-free mMIMO with greedy pilot assignment to that of the DD-mMIMO system. The results indicate that our proposed architecture improves SE. Additionally, we also explore how various parameters affect system performance, including different combining techniques, coordination region radius, total antenna density, and multiple-antenna AP configurations.

E. Outline

The reminder of the paper is organized as follows. In section II, we define the system model, and section III presents the achievable uplink SE and gives other performance analysis. Then, the numerical results and discussions are given in section IV. Finally, we conclude the paper in section V.

Notation: We will adopt the following notations in the rest of the paper. The superscripts $()^*$, $()^T$, and $()^H$ represent the complex conjugate, transpose, and conjugate-transpose, respectively. An uppercase boldface letter stands for a matrix, while a lowercase boldface letter represents a vector. Variables are expressed as italic letters, which are also used for the indices. In contrast, regular (or normal) lowercase letters are used specifically as indices to indicate variable names. The notation $\mathbb{E}\{\cdot\}$ denotes the expectation operator. Finally, $x \sim \mathcal{CN}(\mu, \sigma^2)$ is a complex Gaussian variable with mean μ and variance σ^2.

II. SYSTEM MODEL

In this section, we describe the system model, drawing on the description given in [30] which is also included here for completeness, and making use of similar terminology. We analyze a DD-mMIMO system consisting of M_{coor} APs that are randomly distributed within the coordination region. These APs provide service to K_{serv} UTs located in the service region, as initially introduced in [14]. The network comprises multiple EPUs, two of which are illustrated in Fig. 1. Each EPU is positioned at the center of a coordination region, defined as a circular area with a radius of r_{coor}, where all APs within this region are coordinated and connected to the corresponding EPU. The coordination regions overlap between adjacent EPUs, implying that an AP may be jointly coordinated by multiple EPUs. Under this definition, UTs served by an AP located within this overlapping region are able to transmit their data to any EPU that serves this overlapping coordination region. However, in this paper, we define a polygonal region closest to each EPU as the service region for that EPU, such
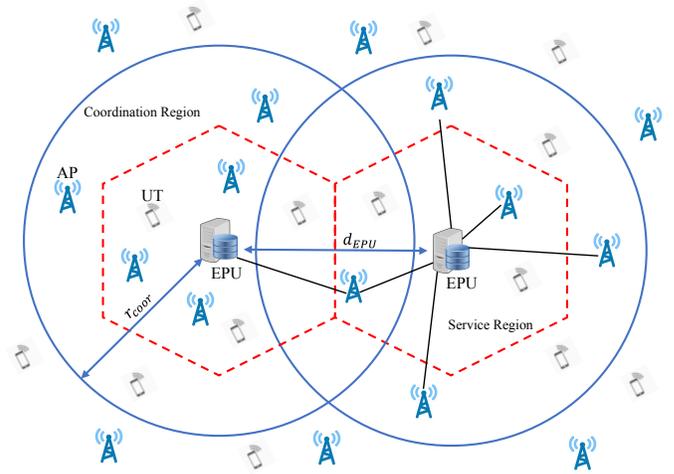

Fig. 1. The architecture of decentralized distributed massive MIMO with circular coordination region and hexagonal service region

that the data from all the UTs in that polygon are processed there.

We assume that the distance between the EPUs, denoted as d_{EPU}, is set to 300 meters, which corresponds to the characteristics of an ultra-dense network where each EPU serves approximately 100 APs, leading to a density exceeding 1000 APs per km^2 [31], [32]. In this paper, we do not delve into the details of the fronthaul connecting the APs to the EPUs; instead, we assume it is error-free and has no bandwidth limitations. To evaluate system performance, we consider a scenario where UTs, each equipped with a single antenna, and multiple-antenna APs are uniformly distributed across the network area with density ρ_{u} and ρ_{A}, respectively. Furthermore, the channel vector between the k-th UT and the m-th AP within the coordination region of a given EPU is denoted as $\mathbf{g}_{mk} = [g_{m1k} \ \ldots \ g_{mN_\text{r}k}]^T$, under the assumption of flat-fading conditions, where N_r represents the number of antennas per AP. It is given by:

$$\mathbf{g}_{mk} = \beta_{mk}^{1/2}\mathbf{h}_{mk} \tag{1}$$

where the vector $\mathbf{h}_{mk} = [h_{m1k} \ \ldots \ h_{mN_\text{r}k}]^T$ represents the small-scale Rayleigh fading, where the elements are independent and identically distributed (i.i.d) complex Gaussian random variables (RVs) following $\mathcal{CN}(0, 1)$. The large-scale fading coefficient β_{mk} is assumed to be identical for the links between a single UT and all N_r antennas at the same AP. Similarly, we define the channel between the k-th interfering UT, located outside the current coordination region, and the m-th AP within the coordination region as $\mathbf{g}_{\text{i},mk} = \beta_{\text{i},mk}^{1/2}\mathbf{h}_{\text{i},mk}$, where $\beta_{\text{i},mk}$ and $\mathbf{h}_{\text{i},mk}$ are the large-scale fading coefficient and Rayleigh fading vector for the interference, respectively.

As in [2], [14], [33], we adopt time-division duplex (TDD) operation for the uplink and downlink transmission. We assume the channel fading is constant over the coherence interval τ_c. Based on the channel reciprocity which is a feature of TDD, the downlink channel knowledge can be acquired through uplink channel estimation, eliminating the need for pilot transmission in the downlink. The coherence interval is

therefore divided into three phases: uplink pilot sequence (τ_{p}), uplink data transmission (τ_{u}), and downlink data transmission (τ_{d}). This relationship can be expressed as $\tau_{\mathrm{c}} = \tau_{\mathrm{p}} + \tau_{\mathrm{u}} + \tau_{\mathrm{d}}$. In this paper, we focus only on the uplink transmission. But, unlike [34], [35] where $\tau_{\mathrm{p}} < K$ (the number of all UTs) in practical cell-free mMIMO, we assume that the pilot length is greater than or equal to the number of UTs in the coordination region, i.e., $\tau_{\mathrm{p}} \geq K_{\mathrm{coor}}$, ensuring that the pilots for all users within each coordination region remain fully orthogonal. The area of the coordination region and service region are given by: $S_{\mathrm{coor}} = \pi r_{\mathrm{coor}}^2$ and $S_{\mathrm{serv}} = \frac{1}{2} \times \frac{d_{\mathrm{EPU}}}{\sqrt{3}} \times \frac{d_{\mathrm{EPU}}}{2} \times 6 = \frac{\sqrt{3}}{2} d_{\mathrm{EPU}}^2$, respectively. Consequently, the average number of UTs in the coordination region can be expressed as $K_{\mathrm{coor}} = \rho_{\mathrm{u}} \pi r_{\mathrm{coor}}^2$, while in the service region, it is given by $K_{\mathrm{serv}} = \frac{\sqrt{3}}{2} \rho_{\mathrm{u}} d_{\mathrm{EPU}}^2$.

III. PERFORMANCE ANALYSIS

A. Channel Estimation

To estimate the channel coefficients, we assume that β_{mk} is known by all APs and EPUs. We suppose that the pilot sequence assigned to the k-th UT within the coordination region is $\boldsymbol{\varphi}_k \in \mathbb{C}^{\tau_{\mathrm{p}} \times 1}$ with $\|\boldsymbol{\varphi}_k\|^2 = \tau_{\mathrm{p}}$. The pilot sequence assigned to the k-th interfering UT outside the coordination region is represented by $\boldsymbol{\varphi}_{\mathrm{i},k}$. As mentioned above, the UTs in the overlapping coordination area are coordinated by both EPUs; therefore, the pilots assigned to these UTs cannot be reused within either of the corresponding coordination regions. However, interfering users outside the coordination region may transmit the same pilots, provided they are not assigned to the UTs within the overlapping area. The received pilot matrix $\mathbf{Y}_{\mathrm{p},m} \in \mathbb{C}^{N_{\mathrm{r}} \times \tau_{\mathrm{p}}}$ at the m-th AP then is given by:

$$\mathbf{Y}_{\mathrm{p},m} = \sum_{k=1}^{K_{\mathrm{coor}}} \mathbf{g}_{mk} \boldsymbol{\varphi}_k^{\mathrm{T}} + \sum_{k=1}^{K_{\mathrm{int}}} \mathbf{g}_{\mathrm{i},mk} \boldsymbol{\varphi}_{\mathrm{i},k}^{\mathrm{T}} + \mathbf{Z}_{\mathrm{p},m} \qquad (2)$$

where K_{int} is the number of interfering UTs, the matrix $\mathbf{Z}_{\mathrm{p},m} \in \mathbb{C}^{N_{\mathrm{r}} \times \tau_{\mathrm{p}}}$ is the noise at the m-th AP, the elements of $\mathbf{Z}_{\mathrm{p},m}$ are i.i.d $\mathcal{CN}(0, \sigma_z^2)$ RVs. For calculating the MMSE estimation, we first take the LS estimate of the channel from the received pilot signals:

$$\check{g}_{mnk} = \frac{1}{\tau_{\mathrm{p}}} \mathbf{y}_{\mathrm{p},mn} \boldsymbol{\varphi}_k^*$$
$$= g_{mnk} + \frac{1}{\tau_{\mathrm{p}}} \sum_{k'=1}^{K_{\mathrm{int}}} g_{\mathrm{i},mnk'} \boldsymbol{\varphi}_{\mathrm{i},k'}^{\mathrm{T}} \boldsymbol{\varphi}_k^* + \frac{1}{\tau_{\mathrm{p}}} \mathbf{z}_{\mathrm{p},mn} \boldsymbol{\varphi}_k^* \quad (3)$$

where $\mathbf{y}_{\mathrm{p},mn}$ and $\mathbf{z}_{p,mn}$ are the n-th row vector of $\mathbf{Y}_{p,m}$ and $\mathbf{Z}_{\mathrm{p},m}$, respectively. Then, we give the MMSE channel estimation:

$$\hat{g}_{mnk} = \frac{\mathbb{E}\{g_{mnk} \check{g}_{mnk}^*\}}{\mathbb{E}\{|\check{g}_{mnk}|^2\}} \check{g}_{mnk} = c_{mnk} \check{g}_{mnk} \qquad (4)$$

where the MMSE weight for the channel estimation is given by:

$$c_{mnk} = \frac{\beta_{mk}}{\beta_{mk} + \frac{1}{\tau_{\mathrm{p}}^2} \sum_{k'=1}^{K_{\mathrm{int}}} \beta_{\mathrm{i},mk'} \mathbb{E}\{|\boldsymbol{\varphi}_{\mathrm{i},k'}^{\mathrm{T}} \boldsymbol{\varphi}_k^*|^2\} + \frac{\sigma_z^2}{\tau_{\mathrm{p}}}} \qquad (5)$$

Here, (5) assumes the terms in (3) are uncorrelated, and we also assume that $\mathbb{E}\{|\boldsymbol{\varphi}_{\mathrm{i},k'}^{\mathrm{T}} \boldsymbol{\varphi}_k^*|^2\} = \tau_{\mathrm{p}}$. Note that c_{mnk} only depends on β_{mk}, hence for all $n = 1, \ldots, N_{\mathrm{r}}$ it remains the same value. So the channel estimates in (4) can be replaced by $\hat{g}_{mnk} = c_{mk} \check{g}_{mnk}$.

The variance of the channel estimation error $\tilde{g}_{mnk}$ for the k-th UT to the n-th antenna of the m-th AP is given by:

$$\begin{aligned} \mathbb{E}\{|\tilde{g}_{mnk}|^2\} &= \mathbb{E}\{|g_{mnk} - \hat{g}_{mnk}|^2\} \\ &= \mathbb{E}\{|g_{mnk}|^2 + |\hat{g}_{mnk}|^2 - g_{mnk} \hat{g}_{mnk}^* \\ &\quad - g_{mnk}^* \hat{g}_{mnk}\} \\ &= \mathbb{E}\{|g_{mnk}|^2 + |\hat{g}_{mnk}|^2 - (\hat{g}_{mnk} + \tilde{g}_{mnk}) \hat{g}_{mnk}^* \\ &\quad - (\hat{g}_{mnk}^* + \tilde{g}_{mnk}^*) \hat{g}_{mnk}\} \\ &\stackrel{(a)}{=} \mathbb{E}\{|g_{mnk}|^2 - |\hat{g}_{mnk}|^2\} \\ &= \beta_{mnk} - c_{mk}^2 \mathbb{E}\{|\check{g}_{mnk}|^2\} \\ &= (1 - c_{mk}) \beta_{mnk} = (1 - c_{mk}) \beta_{mk} \end{aligned} \qquad (6)$$

where step (a) is established because the channel estimates are uncorrelated with their estimation errors.

B. Uplink Transmission and Data Detection

In this section, we discuss the data transmission on the uplink of the DD-mMIMO system. Then, we apply different combining techniques to estimate the data in the EPU and derive an expression for the corresponding SE. We assume τ_{d} binary phase-shift keying (BPSK) data are transmitted by all UTs, and the received signals at all APs can be expressed as a matrix $\mathbf{Y}_{\mathrm{u}} \in \mathbb{C}^{M_{\mathrm{coor}} N_{\mathrm{r}} \times \tau_{\mathrm{d}}}$:

$$\mathbf{Y}_{\mathrm{u}} = \mathbf{G}\mathbf{X} + \mathbf{G}_{\mathrm{int}}\mathbf{X}_{\mathrm{int}} + \mathbf{Z} \qquad (7)$$

where $\mathbf{G} \in \mathbb{C}^{M_{\mathrm{coor}} N_{\mathrm{r}} \times K_{\mathrm{coor}}}$ denotes the channel matrix between the UTs and the APs within the coordination region. $G_{\mathrm{int}} \in \mathbb{C}^{M_{\mathrm{coor}} N_{\mathrm{r}} \times K_{\mathrm{int}}}$ is the channel matrix for the interfering users. The transmitted data, containing τ_{d} symbols, are denoted by the matrix $\mathbf{X} \in \mathbb{R}^{K_{\mathrm{coor}} \times \tau_{\mathrm{d}}}$ and $\mathbf{X}_{int} \in \mathbb{R}^{K_{\mathrm{int}} \times \tau_{\mathrm{d}}}$, within and outside the coordination region, respectively. In this case, we assume that the variance of transmitted data is σ_{x}^2 for all UTs. The elements of the noise matrix $\mathbf{Z}$ follow the distribution $\mathcal{CN}(0, \sigma_z^2)$ and they are uncorrelated between the antennas at the same AP. We also assume all data symbols are independent, so each column vector of the data symbol matrix can be considered independently. Then, we rewrite (7) in a vector form as:

$$\mathbf{y}_{\mathrm{u}} = \sum_{k=1}^{K_{\mathrm{coor}}} \mathbf{g}_k x_k + \sum_{k=1}^{K_{\mathrm{int}}} \mathbf{g}_{\mathrm{i},k} x_{\mathrm{i},k} + \mathbf{z} \qquad (8)$$

where $\mathbf{y}_{\mathrm{u}} \in \mathbb{C}^{M_{\mathrm{coor}} N_{\mathrm{r}} \times 1}$, $\mathbf{g}_k = [\mathbf{g}_{1k}^{\mathrm{T}} \ \cdots \ \mathbf{g}_{M_{\mathrm{coor}} k}^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{g}_{\mathrm{i},k} = [\mathbf{g}_{\mathrm{i},1k}^{\mathrm{T}} \ \cdots \ \mathbf{g}_{\mathrm{i},M_{\mathrm{coor}} k}^{\mathrm{T}}]^{\mathrm{T}}$, and $\mathbf{z}$ is the column vector of $\mathbf{Z}$.

The data is then recovered by weighting the received signals:

$$\text{SINR}_k = \frac{|\mathbf{w}_k \hat{\mathbf{g}}_k|^2}{\mathbf{w}_k \left(\sum_{k' \neq k}^{K_{\text{coor}}} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^{\text{H}} + \sum_{k'=1}^{K_{\text{coor}}} \mathbf{C}_{\tilde{\mathbf{g}},k'} + \sum_{k'=1}^{K_{\text{int}}} \mathbf{C}_{\text{g}_{\text{int}},k'} + \frac{\sigma_z^2}{\sigma_x^2} \mathbf{I}_{M_{\text{coor}} N_r} \right) \mathbf{w}_k^{\text{H}}} \tag{11}$$

$$
\begin{aligned}
\hat{x}_k &= \mathbf{w}_k \mathbf{y}_u \\
&= \underbrace{\mathbf{w}_k \hat{\mathbf{g}}_k x_k}_{\substack{\text{desired signal over estimated channel}}} + \underbrace{\mathbf{w}_k \tilde{\mathbf{g}}_k x_k}_{\substack{\text{desired signal over erroneously} \\ \text{estimated channel}}} \\
&\quad + \underbrace{\sum_{k' \neq k}^{K_{coor}} \mathbf{w}_k \mathbf{g}_{k'} x_{k'}}_{\text{intra-cluster interference}} + \underbrace{\sum_{k'=1}^{K_{int}} \mathbf{w}_k \mathbf{g}_{i,k'} x_{i,k'}}_{\text{inter-cluster interference}} + \underbrace{\mathbf{w}_k \mathbf{z}}_{\text{noise}}
\end{aligned} \tag{9}
$$

where the weight vector $\mathbf{w}_k \in \mathbb{C}^{1 \times M_{coor} N_r}$ can be utilized to express various combining techniques. The term $\hat{\mathbf{g}}_k$, which has the same vector form as $\mathbf{g}_k$, represents the estimated channel vector for the k-th UT to all APs with multiple antennas in the coordination area and $\tilde{\mathbf{g}}_k$ is the channel estimation error vector. In the expanded equality (9), the first two terms correspond to the desired signals where $\mathbf{w}_k \hat{\mathbf{g}}_k x_k$ represents the signal over the known estimated channel, while $\mathbf{w}_k \tilde{\mathbf{g}}_k x_k$ accounts for the component from the erroneously estimated channel. The third term denotes intra-cluster interference originating from within the coordination region, and the fourth term corresponds to inter-cluster interference caused by signals outside the coordination region. The final term is regarded as weighted noise. Then the achievable SE for the k-th UT can be calculated by using the standard capacity bounds, which is shown in the following proposition.

Proposition 1: Consider a DD-mMIMO system where K_{serv} UTs are served by M_{coor} APs, each of which is equipped with N_r antennas. The channel estimation and data detection are performed at the EPU, which is connected to the APs with unlimited fronthaul. An achievable SE for the k-th UT is

$$\text{SE}_k = \frac{\tau_u}{\tau_p + \tau_u} \mathbb{E}\{\log_2(1 + \text{SINR}_k)\} \tag{10}$$

where the instantaneous SINR for the k-th UT (11) is shown at the top of the page and the expectation is with respect to the channel estimates. The factor $\tau_u / (\tau_p + \tau_u)$ is the fraction of uplink transmission which is used for data transmission. The covariance matrices of the channel estimation error for the k-th UT and the channel coefficients for the k-th uncoordinated UT in (11) are expressed as $\mathbf{C}_{\tilde{\mathbf{g}},k}$ and $\mathbf{C}_{\text{g}_{\text{int}},k}$, respectively. $\mathbf{I}_{M_{\text{coor}} N_r} \in \mathbb{R}^{M_{\text{coor}} N_r}$ denotes the identity matrix.

Proof: The proof is similar to the proof in [12, Th. 4.1] and simply proved in Appendix A.

Various combining techniques can be applied to compute the weight vector $\mathbf{w}_k$. A low-complexity combining technique is MRC, where $\mathbf{w}_k = \hat{\mathbf{g}}_k^{\text{H}}$. The numerical results will be discussed in section IV by using Monte Carlo methods. Moreover, we can also use the MMSE estimator to maximize (11) as follows.

Corollary 1: The MMSE combining vector for maximizing the instantaneous SINR in (11) is given by

$$
\begin{aligned}
\mathbf{w}_k &= \mathbb{E}\{|x_k \mathbf{y}_u^{\text{H}}|\} \mathbb{E}\{|\mathbf{y}_u|^2\}^{-1} \\
&= \hat{\mathbf{g}}_k^{\text{H}} \left[\sum_{k'=1}^{K_{\text{coor}}} \left(\hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^{\text{H}} + \mathbf{C}_{\tilde{\mathbf{g}},k'} \right) + \sum_{k'=1}^{K_{\text{int}}} \mathbf{C}_{\text{g}_{\text{int}},k'} + \frac{\sigma_z^2}{\sigma_x^2} \mathbf{I}_{M_{\text{coor}} N_r} \right]^{-1}
\end{aligned} \tag{12}
$$

which leads to the maximum value

$$
\begin{aligned}
\text{SINR}_k = \hat{\mathbf{g}}_k^{\text{H}} &\left(\sum_{k' \neq k}^{K_{\text{coor}}} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^{\text{H}} + \sum_{k'=1}^{K_{\text{coor}}} \mathbf{C}_{\tilde{\mathbf{g}},k'} \right. \\
&\left. + \sum_{k'=1}^{K_{\text{int}}} \mathbf{C}_{\text{g}_{\text{int}},k'} + \frac{\sigma_z^2}{\sigma_x^2} \mathbf{I}_{M_{\text{coor}} N_r} \right)^{-1} \hat{\mathbf{g}}_k \quad (13)
\end{aligned}
$$

Proof: The proof is given in Appendix B.

If the MMSE estimator is used to estimate channels for all UTs in the coordination region, the covariance matrix of channel estimation errors in (13) is then given by:

$$
\begin{aligned}
\mathbf{C}_{\tilde{\mathbf{g}},k} &= \mathbb{E}\{\tilde{\mathbf{g}}_k \tilde{\mathbf{g}}_k^{\text{H}}\} \\
&= \text{diag}\{ \underbrace{(1 - c_{1k})\beta_{1k} \ldots (1 - c_{1k})\beta_{1k}}_{N_r}, \ldots \ldots, \\
&\quad \underbrace{(1 - c_{M_{\text{coor}}k})\beta_{M_{\text{coor}}k} \ldots (1 - c_{M_{\text{coor}}k})\beta_{M_{\text{coor}}k}}_{N_r} \} \quad (14)
\end{aligned}
$$

where diag denotes diagonal matrix containing non-zero elements only along its main diagonal. The on-diagonal elements are calculated using (6). Since the channels among antennas at the same AP are assumed uncorrelated, the off-diagonal elements in (14) are zero.

The covariance matrix of interfering channels in (13) is:

$$
\begin{aligned}
\mathbf{C}_{\text{g}_{\text{int}},k} &= \mathbb{E}\{\tilde{\mathbf{g}}_{i,k} \tilde{\mathbf{g}}_{i,k}^{\text{H}}\} \\
&= \text{diag}\{ \underbrace{\beta_{i,1k} \ldots \beta_{i,1k}}_{N_r}, \ldots \ldots, \underbrace{\beta_{i,M_{\text{coor}}k} \ldots \beta_{i,M_{\text{coor}}k}}_{N_r} \} \quad (15)
\end{aligned}
$$

C. Scalability

As discussed above, we know that scalability is a very important feature for the Cell-free network. Based on the definition of scalability given in [24], we add some comments to this definition which clearly explain why our system fulfills the demand of finite complexity. In the DD-mMIMO system, all APs within the coordination region transmit the received pilots and data to the EPU to provide the channel state information (CSI) and hence recover the data. The processing does not occur at the AP. Hence, a finite complexity is not only required for each AP but also for each EPU in the original definition in [24]. In detail, each EPU estimates the channels

only for the M_{coor} APs associated with K_{coor} UTs within the coordination region. Consequently, the channel estimation complexity for each EPU remains finite, even as the total number of UTs in the network K tends to infinity. In order to detect the data sent by all UTs within the service region, the EPU calculates $\{\mathbf{w}_k \mathbf{y}_{\text{u}} : k = 1, \ldots, K_{\text{serv}}\}$ in (9) using K_{serv} combining vectors $\{\mathbf{w}_k : k = 1, \ldots, K_{\text{serv}}\}$. This process also exhibits finite computational complexity, regardless of the combining techniques used. It is important to note that when using MMSE combining vectors, as described in (12), the covariance matrices of interfering channels $\{\mathbf{C}_{\text{g}_{\text{int}},k} : k = 1, \ldots, K_{\text{int}}\}$ are aggregated with respect to the total number of interfering users, K_{int}. While this complexity could grow unboundedly as K approaches infinity, the covariance matrix $\mathbf{C}_{\text{g}_{\text{int}},k}$ in (15) depends only on the large scale fading coefficients, which are determined by the UT-AP distance. For UTs that are sufficiently distant, these coefficients become negligible. Therefore, by setting a threshold for the UT-AP distance and disregarding any UTs beyond this limit, the overall computational complexity is ensured to remain finite.

During the uplink, each AP within the coordination region forwards received pilots of length τ_{p} and data signals of length τ_{u} over the fronthaul link to an EPU. In the downlink transmission, each EPU generates transmitted signals for M_{coor} APs within the coordination region, which include the precoding vectors and data signals intended for K_{coor} UTs. Each AP then receives the specific downlink data signals designated for the K_{coor} UTs. As a result, the fronthaul signaling load remains finite. Finally, this study does not implement power control in the DD-mMIMO system; however, it demonstrates that DD-mMIMO still achieves superior performance compared to cell-free mMIMO. Thus, the fourth task defined in [24, Definition 1] is deemed unnecessary at the current stage but remains a potential area for future research. Overall, the DD-mMIMO system is demonstrated to be scalable.

IV. NUMERICAL RESULTS AND DISCUSSION

In this section, we first present the large-scale fading model and set up the simulation parameters. We then provide numerical results comparing our proposed DD-mMIMO system with the canonical cell-free mMIMO system using greedy pilot allocation to evaluate the achievable uplink SE. We also examine the system performance with different receive combining techniques and varying the radius of the coordination region. Additionally, we consider a scenario with the same antenna density but different numbers of antennas per AP. The results for multiple-antenna AP are discussed thereafter.

A. Large-Scale Fading Model

The large-scale fading coefficient in (1) is modeled by the path loss and the shadow fading:

$$\beta_{mk} = 10^{\frac{-\text{PL}_{mk} + \text{F}_{mk}}{10}} \quad (16)$$

where PL_{mk} denotes the path loss and $\text{F}_{mk} = \sigma_{\text{F}}^2 \alpha_{mk}$ represents the shadow fading with standard deviation σ_{F} and $\alpha_{mk} \sim \mathcal{N}(0,1)$. In this paper, we consider only uncorrelated
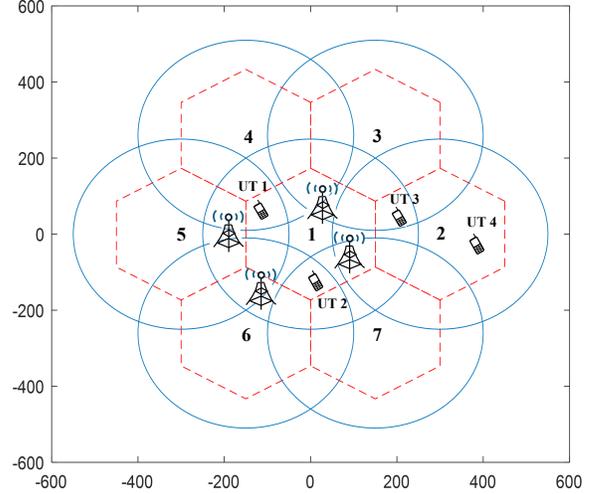

Fig. 2. The simulation area for the architecture of DD-mMIMO with six neighboring EPUs and their service and coordination regions.

shadow fading. To give a comprehensive analysis, we adopt the path loss model for the Non-Line of Sight (NLOS) urban microcell (UMi) scenario given in the 3GPP standard [36]. The path loss in dB form is expressed by:

$$\text{PL}_{mk} = 35.3\log_{10}(d_{\text{3D}}) + 22.4 + 21.3\log_{10}(f_{\text{c}}) \\ - 0.3(h_{\text{UT}} - 1.5) \quad (17)$$

where d_{3D} is the distance in three dimensions in meters between the antenna of the AP and the UT, f_c represents the carrier frequency in GHz, and h_{UT} denotes the height of the UT.

B. Simulation Setup

A DD-mMIMO system is considered, featuring a hexagonal service area and a circular coordination area. The inter-EPU distance is set to 300 m, and the entire system is simulated within a square area measuring 1 km on each side. To give a comprehensive analysis, the neighboring EPUs with their coordination regions are considered, as shown in Fig. 2. Given that the UTs located inside the service region are served by their own EPU, while those positioned outside the coordination region generate interference, each EPU with its service region and its coordination region can be evaluated individually. In this paper, we mainly focus on the UT's performance in the central service area.

We consider a UT density $\rho_{\text{u}} = 40/\text{km}^2$ and an AP density $\rho_{\text{A}} = 160/\text{km}^2$ unless otherwise stated. The path loss model is defined by (17), using the following parameters: $f_{\text{c}} = 1.9$ GHz, the AP height $h_{\text{AP}} = 10$m and $h_{\text{UT}} = 1.65$m. All UTs transmit with power $\sigma_{\text{x}}^2 = \sigma_{\text{i,x}}^2 = 100$ mW, while the noise power is configured as $\sigma_{\text{z}}^2 = -96$ dBm. The system operates with a bandwidth 20 MHz. The uplink transmission duration is set to $\tau_{\text{u}} = 190$ data symbols per packet, with a pilot length of $\tau_{\text{p}} = 10$. Additionally, the locations of UTs and

APs are randomly distributed within the simulation area, with an inter-AP distance assumed to be greater than 10 meters. The locations of UTs and APs included in the simulation are limited by the geometric boundary. To give a realistic simulation, the number of UTs and APs follows the random Poisson distribution with parameters $\lambda = \rho_u \times 1\text{km}^2$ and $\lambda = \rho_A \times 1\text{km}^2$, respectively.

For pilot allocation, we assume the simulated system consists of a central EPU and its six surrounding EPUs, as shown in Fig. 2, with overlapping portions among the coordination regions. As stated in Section III, the pilots assigned to an overlapping area cannot be reused by other UTs located within any of the overlapping coordination regions. The UTs outside these seven EPUs are located farther away, resulting in relatively lower interference due to smaller path loss, which does not significantly impact the desired signal power. Therefore, we assume that these UTs are randomly assigned pilot sequences. To illustrate this pilot allocation method, we give a simple example as follows:

Example 1. *In Fig. 2, we first assume a table is created to record the used and unused pilot indices in each EPU, and this table is shared among all neighboring EPUs (labeled EPU1 to EPU7 with black numbers 1 to 7). Four UTs (UT1 to UT4) are successively added to the network. The first three UTs are assigned pilots chosen from the pilot sequences φ, ensuring mutual orthogonality. The used pilot indices are then recorded in each relevant table; for example, the pilot index used by UT1 is recorded in the tables of EPU1, EPU4 and EPU5, respectively. Finally, UT4 selects a pilot index from the table, excluding the one used by UT3.*

C. Numerical Results

Fig. 3 shows the cumulative distribution function (CDF) of achievable uplink SE for both the cell-free mMIMO system and the DD-mMIMO system, with MRC and MMSE data detection. Note that MMSE channel estimation is used in both systems and that greedy pilot assignment is applied for cell-free mMIMO. To ensure a fair comparison, the radius of the coordination region for the DD-mMIMO system is set to 282 meters. This is based on the assumption in this paper that the pilot length is 10 and the density of UTs is 40. As a result, the entire simulation area, modeled as a cell-free mMIMO network, consists of 10 UTs using fully orthogonal pilots, while the remaining 30 UTs are treated as sources of interference due to pilot contamination. With a coordination region radius of 282 meters, the average number of UTs within the coordination region is approximately 10, ensuring that both systems experience a similar level of useful signal and interference. Note that while cell-free mMIMO processes signals from all AP-UT pairs across the entire simulation area, only the results for UTs within the service region are used in comparisons with DD-mMIMO.

The results indicate that for 95% of UTs, the SE of DD-mMIMO improves by 18.5% compared to cell-free mMIMO using MMSE data detection, and by 27.7% at the median SE point. In the case of MRC data detection, both systems exhibit similar SE performance; however, cell-free mMIMO performs slightly better for 20% of the UTs, as observed at the lower end of the curves. Moreover, for 95% of UTs, the SE achieved using the MMSE combining technique is nearly 7.5 times higher than that with MRC in DD-mMIMO. In (12), calculating the MMSE weight for DD-mMIMO requires inverting an $M_{\text{coor}}N_r \times M_{\text{coor}}N_r$ matrix, with a computational complexity of $\mathcal{O}(M_{\text{coor}}^3 N_r^3)$, where $\mathcal{O}(\cdot)$ denotes big O notation. In contrast, cell-free mMIMO requires inverting an $MN_r \times MN_r$ matrix within a single centralized CPU, resulting in a significantly higher computational complexity compared to DD-mMIMO.
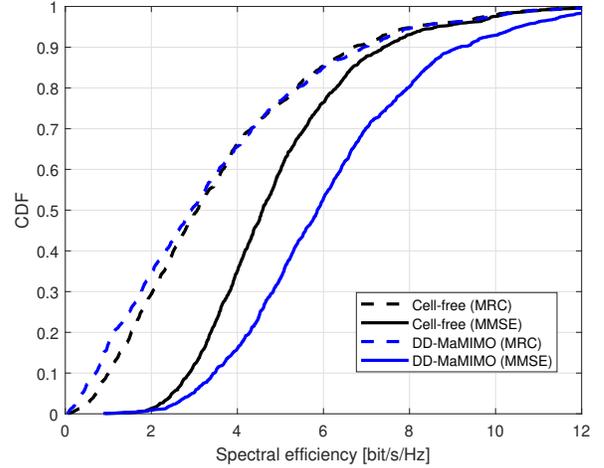

Fig. 3. The cumulative distribution function of uplink SE for the comparison between cell-free mMIMO (black line) and DD-mMIMO (blue line) with MRC (dashed line) and MMSE (solid line) data detection.
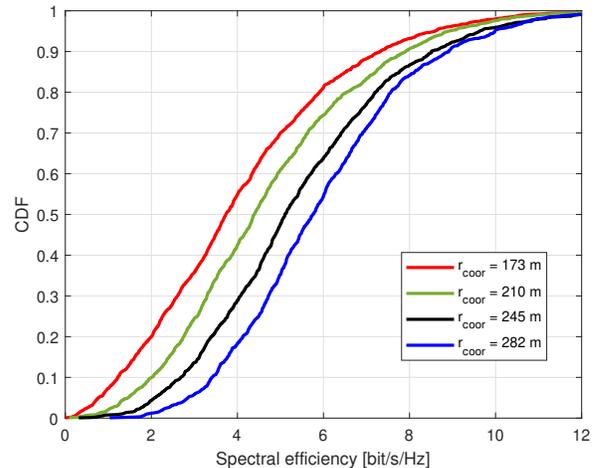

Fig. 4. The CDF of SE for DD-mMIMO with varying radii of coordination region.

Fig. 4 presents the CDF of SE for DD-mMIMO with varying coordination region radii, illustrating that SE progressively increases by 18%, 16.1%, and 12.7% as the radius expands from 173 meters to 282 meters, with each 35-meter increment. The proportion of improvement diminishes, indicating a limitation in SE. However, this limitation is beyond the scope of this paper and will be explored in future work.

We adopt a minimum radius of 173 meters, as this distance allows the coordination region to fully cover the service region. In other words, this scenario is equivalent to a cell-free mMIMO system divided into many clusters, each served by an independent CPU. The results show that the 173-meter radius case exhibits poor performance due to edge effects, where the UTs near the cluster edge can cause strong interference.

Then, Fig. 5 presents two scenarios with the same total antenna density of 160/km^2, but with varying numbers of antennas per AP and different numbers of APs. The scenario with more APs, each equipped with a single antenna, achieves better SE performance than the case with fewer APs equipped with multiple antennas. This is because with fewer APs the average distance between an AP and a UT increases, leading to higher path loss. Note that the radius of the coordination region is 282 meters in this case.

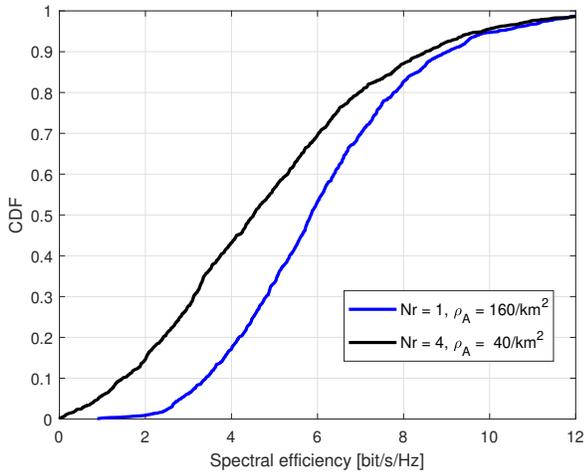Fig. 6. The SE for different numbers of antennas per AP, considering both 95% of UTs and median UTs.

Fig. 5. The CDF of SE for DD-mMIMO with the same antenna density but varying numbers of antennas per AP.

The SE for multiple antennas per AP is presented in Fig. 6, where a 282-meter coordination region radius is adopted. For 95% of UTs, the SE reaches 6 bits/s/Hz with the 8-antenna APs, which is twice that of the case with APs equipped with a single antenna. Meanwhile, the SE for the median UTs shows a significant improvement, increasing from 5.8 bit/s/Hz to 9.3 bit/s/Hz.

V. Conclusion

In this paper, we have explored DD-mMIMO, which is a promising architecture to fulfill the ultra-dense and low-latency requirement for 6G and beyond. We compared our proposed architecture with the cell-free mMIMO using greedy pilot allocation, demonstrating that DD-mMIMO with MMSE data detection significantly improves the SE. Additionally, we derived general expressions for the DD-mMIMO system to calculate the uplink SINR and SE, which are suitable for different combining techniques. Numerical results showed that using MMSE combining outperforms MRC in terms of SE. Moreover, DD-mMIMO is scalable and offers good performance in general scenarios without relying on the "channel
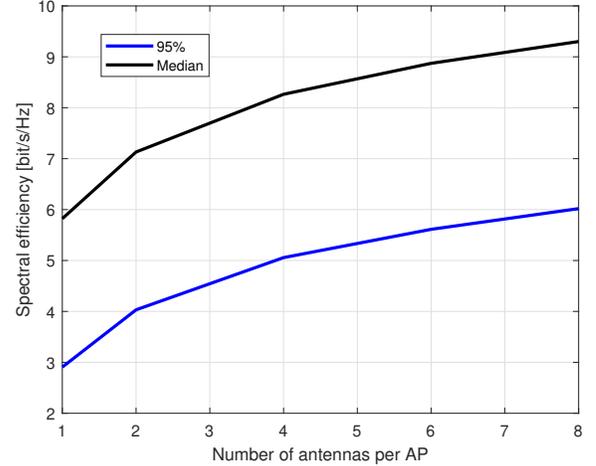
hardening" assumption. Finally, we evaluated our proposed architecture with varying parameters, such as the radius of the coordination region, the density for the total antenna density, and the number of antennas per AP.

APPENDIX A
PROOF OF PROPOSITION 1

For simply calculating the instantaneous SINR, we firstly introduce the term v which includes all interference and noise in (9) as follows:

$$v = \mathbf{w}_k\tilde{\mathbf{g}}_k x_k + \sum_{k'\neq k}^{K_{\text{coor}}} \mathbf{w}_k\mathbf{g}_{k'} x_{k'} + \sum_{k'=1}^{K_{\text{int}}} \mathbf{w}_k\mathbf{g}_{\text{i},k'} x_{\text{i},k'} + \mathbf{w}_k\mathbf{z} \quad (18)$$

where the terms in v are independent of each other, and both the data x and noise z have a zero mean. Thus, v has a zero mean. To obtain the interference plus noise power, we calculate the conditional variance of v, which is affected by the random realization $\hat{\mathbf{g}}_k$:

$$\mathbb{E}\{|v|^2|\{\hat{\mathbf{g}}_k\}\} \stackrel{(a)}{=} \mathbb{E}\{|x_k|^2\}\mathbb{E}\{|\mathbf{w}_k\tilde{\mathbf{g}}_k|^2|\{\hat{\mathbf{g}}_k\}\}$$

$$+ \sum_{k'\neq k}^{K_{\text{coor}}} \mathbb{E}\{|x_{k'}|^2\}\mathbb{E}\{|\mathbf{w}_k\tilde{\mathbf{g}}_{k'}|^2|\{\hat{\mathbf{g}}_k\}\}$$

$$+ \sum_{k'=1}^{K_{\text{int}}} \mathbb{E}\{|x_{\text{i},k'}|^2\}\mathbb{E}\{|\mathbf{w}_k\tilde{\mathbf{g}}_{\text{i},k'}|^2|\{\hat{\mathbf{g}}_k\}\}$$

$$+ \mathbb{E}\{|\mathbf{w}_k\mathbf{z}|^2|\{\hat{\mathbf{g}}_k\}\}$$

$$\stackrel{(b)}{=} \sigma_{\text{x}}^2\mathbf{w}_k\mathbf{C}_{\tilde{\text{g}},k}\mathbf{w}_k^{\text{H}}$$

$$+ \sum_{k'\neq k}^{K_{\text{coor}}} \sigma_{\text{x}}^2\mathbf{w}_k(\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^{\text{H}} + \mathbf{C}_{\tilde{\text{g}},k'})\mathbf{w}_k^{\text{H}} \qquad (19)$$

$$+ \sum_{k'=1}^{K_{\text{int}}} \sigma_{\text{x}}^2\mathbf{w}_k\mathbf{C}_{\text{g}_{\text{int}},k'}\mathbf{w}_k^{\text{H}} + \mathbf{w}_k\sigma_{\text{z}}^2\mathbf{I}\mathbf{w}_k^{\text{H}}$$

$$= \sigma_{\text{x}}^2\mathbf{w}_k\left(\sum_{k'\neq k}^{K_{\text{coor}}}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^{\text{H}} + \sum_{k'=1}^{K_{\text{coor}}}\mathbf{C}_{\tilde{\text{g}},k'}\right.$$

$$\left. + \sum_{k'=1}^{K_{\text{int}}}\mathbf{C}_{\text{g}_{\text{int}},k'} + \frac{\sigma_{\text{z}}^2}{\sigma_{\text{x}}^2}\mathbf{I}\right)\mathbf{w}_k^{\text{H}}$$

where step (a) is established because the input signal x is uncorrelated with v given the channel estimates. For example, $\mathbb{E}\{x_k v|\{\hat{\mathbf{g}}_k\}\} = \mathbb{E}\{\mathbf{w}_k\tilde{\mathbf{g}}_k|\{\hat{\mathbf{g}}_k\}\}\mathbb{E}\{|x_k|^2\} = 0$ in which the channel estimation error is zero mean. Step (b) follows from the fact that the channel estimates are uncorrelated with their corresponding errors.

Next, the desired signal power is calculated by $\mathbb{E}\{|\mathbf{w}_k\hat{\mathbf{g}}_k x_k|^2|\{\hat{\mathbf{g}}_k\}\} = \sigma_x^2|\mathbf{w}_k\hat{\mathbf{g}}_k|^2$. Finally, (11) is proved, and the capacity bound provided in [12, Corollary 1.3] can be utilized to compute SE.

Appendix B
Proof of Corollary 1

First, we use $\mathbf{D}$ to represent the denominator of (11) and rewrite it as

$$\text{SINR}_k = \frac{|\mathbf{w}_k\hat{\mathbf{g}}_k|^2}{\mathbf{w}_k\mathbf{D}\mathbf{w}_k^{\text{H}}} \qquad (20)$$

The MMSE weight vector (12) can also be expressed as $\mathbf{w}_k = \hat{\mathbf{g}}_k^{\text{H}}(\mathbf{D} + \hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1}$. Then, we substitute this weight into (20) and simplify it as

$$\text{SINR}_k = \frac{\mathbf{w}_k\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}\mathbf{w}_k^{\text{H}}}{\mathbf{w}_k\mathbf{D}\mathbf{w}_k^{\text{H}}}$$

$$\stackrel{(a)}{=} \frac{\hat{\mathbf{g}}_k^{\text{H}}(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1}\hat{\mathbf{g}}_k}{\hat{\mathbf{g}}_k^{\text{H}}(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1}\mathbf{D}(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1}\hat{\mathbf{g}}_k}$$

$$\stackrel{(b)}{=} \frac{\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k}{\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\mathbf{D}\mathbf{D}^{-1}\hat{\mathbf{g}}_k} = \hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k \qquad (21)$$

$$= \hat{\mathbf{g}}_k^{\text{H}}\left(\sum_{k'\neq k}^{K_{\text{coor}}}\hat{\mathbf{g}}_{k'}\hat{\mathbf{g}}_{k'}^{\text{H}} + \sum_{k'=1}^{K_{\text{coor}}}\mathbf{C}_{\tilde{\text{g}},k'} + \sum_{k'=1}^{K_{\text{int}}}\mathbf{C}_{\text{g}_{\text{int}},k'}\right.$$

$$\left. + \frac{\sigma_{\text{z}}^2}{\sigma_{\text{x}}^2}\mathbf{I}\right)^{-1}\hat{\mathbf{g}}_k$$

where step (a) is obtained because of $\mathbf{D} = \mathbf{D}^{\text{H}}$. Step (b) is established by applying the Woodbury formula: $(A+UCV)^{-1} = A^{-1} - A^{-1}U(VA^{-1}U + C^{-1})^{-1}VA^{-1}$ and here, we assume $A = \mathbf{D}$, $U = \hat{\mathbf{g}}_k$, $C = 1$ and $V = \hat{\mathbf{g}}_k^{\text{H}}$. Consequently, the resulting expression is given as:

$$(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1} = \mathbf{D}^{-1} - \frac{1}{1+\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k}\mathbf{D}^{-1}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1} \quad (22)$$

where the result of $\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k$ is a scalar. Furthermore, we calculate

$$\hat{\mathbf{g}}_k^{\text{H}}(\mathbf{D}+\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}})^{-1} = \hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1} - \frac{\hat{\mathbf{g}}_k^{\text{H}}}{1+\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k}\mathbf{D}^{-1}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}$$

$$= \frac{(1+\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k)\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1} - \hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}}{1+\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k}$$

$$= \frac{\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}}{1+\hat{\mathbf{g}}_k^{\text{H}}\mathbf{D}^{-1}\hat{\mathbf{g}}_k}$$

$$(23)$$

Then, the denominator of (23) is canceled in the equality of (21). The corollary is proved.

References

[1] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE transactions on wireless communications*, vol. 9, no. 11, pp. 3590–3600, 2010.

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

[3] S. Elhoushy, M. Ibrahim, and W. Hamouda, "Cell-free massive MIMO: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 492–523, 2021.

[4] A. Burr, S. Islam, J. Zhao, and M. Bashar, "Cell-free massive MIMO with multi-antenna access points and user terminals," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020, pp. 821–825.

[5] E. Björnson and L. Sanguinetti, "Making cell-free massive MIMO competitive with MMSE processing and centralized implementation," *IEEE Transactions on Wireless Communications*, vol. 19, no. 1, pp. 77–90, 2019.

[6] S. Willhammar, J. Flordelis, L. Van der Perre, and F. Tufvesson, "Channel hardening in massive MIMO-A measurement based analysis," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2018, pp. 1–5.

[7] M. Matthaiou, O. Yurduseven, H. Q. Ngo, D. Morales-Jimenez, S. L. Cotton, and V. F. Fusco, "The road to 6G: Ten physical layer challenges for communications engineers," *IEEE Communications Magazine*, vol. 59, no. 1, pp. 64–69, 2021.

[8] W. Roh and A. Paulraj, "Outage performance of the distributed antenna systems in a composite fading channel," in *Proceedings IEEE 56th Vehicular Technology Conference*, vol. 3. IEEE, 2002, pp. 1520–1524.

[9] D. Castanheira and A. Gameiro, "Distributed antenna system capacity scaling [coordinated and distributed mimo]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 68–75, 2010.

[10] E. Björnson, M. Matthaiou, A. Pitarokoilis, and E. G. Larsson, "Distributed massive MIMO in cellular networks: Impact of imperfect hardware and number of oscillators," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. IEEE, 2015, pp. 2436–2440.

[11] T. L. Marzetta, E. G. Larsson, H. Yang, and H. Q. Ngo, *Fundamentals of massive MIMO*. Cambridge University Press, 2016.

[12] E. Björnson, J. Hoydis, L. Sanguinetti *et al.*, "Massive MIMO networks: Spectral, energy, and hardware efficiency," *Foundations and Trends® in Signal Processing*, vol. 11, no. 3-4, pp. 154–655, 2017.

[13] 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 Transactions on Wireless Communications*, vol. 18, no. 4, pp. 2021–2036, 2019.

[14] A. Burr, M. Bashar, and D. Maryopi, "Ultra-dense radio access networks for smart cities: Cloud-RAN, fog-RAN and" cell-free" massive MIMO," *arXiv preprint arXiv:1811.11077*, 2018.

[15] C. Mobile, "C-RAN: the road towards green RAN," *White paper, ver*, vol. 2, no. 5, pp. 15–16, 2011.

[16] 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 Communications surveys & tutorials*, vol. 17, no. 1, pp. 405–426, 2014.

[17] M. A. Habibi, M. Nasimi, B. Han, and H. D. Schotten, "A comprehensive survey of RAN architectures toward 5g mobile communication system," *Ieee Access*, vol. 7, pp. 70 371–70 421, 2019.

[18] Y.-Y. Shih, W.-H. Chung, A.-C. Pang, T.-C. Chiu, and H.-Y. Wei, "Enabling low-latency applications in fog-radio access networks," *IEEE network*, vol. 31, no. 1, pp. 52–58, 2016.

[19] Y.-J. Ku, D.-Y. Lin, C.-F. Lee, P.-J. Hsieh, H.-Y. Wei, C.-T. Chou, and A.-C. Pang, "5G radio access network design with the fog paradigm: Confluence of communications and computing," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 46–52, 2017.

[20] xRAN Forum, "xRAN Forum Merges With C-RAN Alliance to Form ORAN Alliance," 2018, [Online]. Available: urlhttps://www.businesswire.com/news/home/20180227005673/en/.

[21] M. Polese, L. Bonati, S. D'oro, S. Basagni, and T. Melodia, "Understanding O-RAN: Architecture, interfaces, algorithms, security, and research challenges," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 1376–1411, 2023.

[22] "Study on new radio access technology: Radio access architecture and interfaces," 3rd Generation Partnership Project (3GPP), Technical Specification TS 38.801, Apr. 2017.

[23] L. Bonati, M. Polese, S. D'Oro, S. Basagni, and T. Melodia, "Open, programmable, and virtualized 5G networks: State-of-the-art and the road ahead," *Computer Networks*, vol. 182, p. 107516, 2020.

[24] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Transactions on Communications*, vol. 68, no. 7, pp. 4247–4261, 2020.

[25] G. Interdonato, P. Frenger, and E. G. Larsson, "Scalability aspects of cell-free massive MIMO," in *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. IEEE, 2019, pp. 1–6.

[26] O. Y. Bursalioglu, G. Caire, R. K. Mungara, H. C. Papadopoulos, and C. Wang, "Fog massive MIMO: A user-centric seamless hot-spot architecture," *IEEE Transactions on Wireless Communications*, vol. 18, no. 1, pp. 559–574, 2018.

[27] N. Rajatheva, I. Atzeni, E. Bjornson, A. Bourdoux, S. Buzzi, J.-B. Dore, S. Erkucuk, M. Fuentes, K. Guan, Y. Hu *et al.*, "White paper on broadband connectivity in 6G," *arXiv preprint arXiv:2004.14247*, 2020.

[28] H. Q. Ngo, G. Interdonato, E. G. Larsson, G. Caire, and J. G. Andrews, "Ultradense cell-free massive MIMO for 6G: Technical overview and open questions," *Proceedings of the IEEE*, 2024.

[29] Z. Chen and E. Björnson, "Can we rely on channel hardening in cell-free massive MIMO?" in *2017 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2017, pp. 1–6.

[30] J. Zhao, "Decentralised distributed massive mimo," Ph.D. dissertation, University of York, 2023.

[31] M. Ding, D. López-Pérez, G. Mao, P. Wang, and Z. Lin, "Will the area spectral efficiency monotonically grow as small cells go dense?" in *2015 IEEE Global Communications Conference (GLOBECOM)*. IEEE, 2015, pp. 1–7.

[32] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Communications surveys & tutorials*, vol. 18, no. 4, pp. 2522–2545, 2016.

[33] G. Interdonato, E. Björnson, H. Quoc Ngo, P. Frenger, and E. G. Larsson, "Ubiquitous cell-free massive MIMO communications," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–13, 2019.

[34] 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 Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 25–39, 2017.

[35] M. Bashar, H. Q. Ngo, K. Cumanan, A. G. Burr, P. Xiao, E. Björnson, and E. G. Larsson, "Uplink spectral and energy efficiency of cell-free massive MIMO with optimal uniform quantization," *IEEE Transactions on Communications*, vol. 69, no. 1, pp. 223–245, 2020.

[36] "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Technical Report TR 38.901, 2022, release 17.

Junbo Zhao received the B.Eng. degree in Electronic Information Engineering from the Beijing Information Science and Technology University, Beijing, China in 2016, and the M.Sc. degree in Communications Engineering and Ph.D. degree in Electronic Engineering from the University of York, U.K. in 2019 and 2023, respectively. He is currently a post-doctoral researcher in the School of Physics, Engineering and Technology, University of York, U.K. His research interests include MIMO and Cell-free massive MIMO, optimal pilot assignment, channel estimation, physical layer data detection and quantization, and Open RAN.

Mostafa Rahmani was born in Iran in 1986. He received his B.Sc. degree in Electrical Engineering from Shiraz University, Iran, in 2009, followed by an M.Sc. degree in Communication Systems Engineering from Tarbiat Modares University, Tehran, Iran, in 2012. He completed his Ph.D. in Communication Systems Engineering at Shiraz University of Technology, Iran, in 2023. Currently, he is a postdoctoral researcher at the University of York. His research interests focus on wireless communication systems, including MIMO and cell-free massive MIMO, Open RAN, physical layer network coding, and the application of machine learning and deep reinforcement learning in communications.

Alister Burr was born in London, U.K, in 1957. He received the BSc degree in Electronic Engineering from the University of Southampton, U.K in 1979 and the PhD from the University of Bristol in 1984. Between 1975 and 1985 he worked at Thorn-EMI Central Research Laboratories in London. In 1985 he joined the Department of Electronics (now part of the School of Physics, Engineering and Technology) at the University of York, U.K, where he has been Professor of Communications since 2000. His research interests are in wireless communication systems, including MIMO and cell-free massive MIMO, Open RAN, physical layer network coding, and iterative detection and decoding techniques. He has published more than 300 papers in refereed international conferences and journals, and is the author of "Modulation and Coding for Wireless Communications" (published by Prentice-Hall/PHEI), and co-author of "Wireless Physical-Layer Network Coding" (Cambridge University Press, 2018). In 1999 he was awarded a Senior Research Fellowship by the U.K. Royal Society, and in 2002 he received the J. Langham Thompson Premium from the Institution of Electrical Engineers. He has also given more than 20 invited presentations, including six keynote presentations. He has been co-chair, working group 2, of a series of European COST programmes including currently CA20120 INTERACT, and has also served as Associate Editor for IEEE Communications Letters, Workshops Chair for IEEE ICC 2016, and TPC co-chair for PIMRC 2018 and 2020.