

Resource Allocation for Cell-free Massive MIMO-enabled URLLC Downlink Systems

Qihao Peng, Hong Ren, Cunhua Pan, Nan Liu, and Maged ElKashlan

Abstract

Ultra-reliable and low-latency communication (URLLC) is a pivotal technique for enabling the wireless control over industrial Internet-of-Things (IIoT) devices. By deploying distributed access points (APs), cell-free massive multiple-input and multiple-output (CF mMIMO) has great potential to provide URLLC services for IIoT devices. In this paper, we investigate CF mMIMO-enabled URLLC in a smart factory. Lower bounds (LBs) of downlink ergodic data rate under finite channel blocklength (FCBL) with imperfect channel state information (CSI) are derived for maximum-ratio transmission (MRT), full-pilot zero-forcing (FZF), and local zero-forcing (LZF) precoding schemes. Meanwhile, the weighted sum rate is maximized by jointly optimizing the pilot power and transmission power based on the derived LBs. Specifically, we first provide the globally optimal solution of the pilot power, and then introduce some approximations to transform the original problems into a series of subproblems, which can be expressed in a geometric programming (GP) form that can be readily solved. Finally, an iterative algorithm is proposed to optimize the power allocation based on various precoding schemes. Simulation results demonstrate that the proposed algorithm is superior to the existing algorithms, and that the quality of URLLC services will benefit by deploying more APs, except for the FZF precoding scheme.

Index Terms

Cell-free massive MIMO, URLLC, Industrial Internet-of-Things (IIoT).

I. INTRODUCTION

Ultra-Reliable Low-Latency Communication (URLLC) is one of the crucial techniques in the next generation industrial systems, which can support the mission-critical communication for industrial Internet-of-Things (IIoT) devices such as autonomous vehicles and robots [1], [2]. For industrial applications, the control command data packet size is generally small with the stringent requirements of low latency (1 ms) and low block error rate below 10^{-6} [3]. Since the blocklength no longer tends to be infinite, the impact of the decoding error probability (DEP) should be considered. To investigate the coding rate in the short packet regime, the authors of [4] derived the

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Manuscript received May 31, 2022; revised October 11, 2022; accepted January 30, 2023. The work of Q. Peng was supported by the China Scholarship Council. This work was supported in part by the National Natural Science Foundation of China (Grant No. 62101128), National Key Research and Development Project (Grant No. 2019YFE0123600), National Natural Science Foundation of China (Grant No. 62201137), and Basic Research Project of Jiangsu Provincial Department of Science and Technology (Grant No. BK20210205).

Q. Peng and M. ElKashlan are with the School of Electronic Engineering and Computer Science at Queen Mary University of London, U.K. (e-mail: {q.peng, maged.elkashlan}@qmul.ac.uk). H. Ren, C. Pan, and N. Liu are with National Mobile Communications Research Laboratory, Southeast University, Nanjing, China. (e-mail: {hren, cpan, nanliu}@seu.edu.cn). (*Corresponding author: Hong Ren, Cunhua Pan.*)

approximated achievable data rate under finite channel blocklength (FCBL), which was expressed in a complex function of the channel blocklength and DEP [5]. However, the achievable data rate expression is neither convex nor concave with respect to channel blocklength and signal-to-noise ratio (SNR) [6], which is challenging for resource allocation.

Recently, there are some contributions on resource allocation based on short packet transmission [7]–[9]. By deploying an unmanned aerial vehicle (UAV) as a relay, the short packet can be delivered to an obstructed device by optimizing UAV’s location and channel blocklength [7]. The joint optimization on power allocation and blocklength was studied in [8]. The overall DEP was minimized by optimizing the power allocation in non-orthogonal multiple access (NOMA) systems [9]. However, all the above studies [7]–[9] only considered a simple scenario with point-to-point link, while a smart industry needs to provide URLLC services for a large number of devices [10]. To support multiple devices, the orthogonal frequency division multiple access (OFDMA) technique was adopted in [11], and the authors therein aimed to minimize the total bandwidth by optimizing the subchannel allocation. However, the frequency resource in IIoT applications is limited [12] and the OFDMA technique is not effective for supporting an excessive number of devices.

Owing to a large number of available spatial degrees of freedom, massive multiple-input and multiple-output (mMIMO) can simultaneously support multiple devices by using the same time-frequency resources [13], [14], and thus the mMIMO-enabled URLLC has attracted extensive research attention [15]–[19]. The pilot length was optimized to minimize the DEP in [15], and the authors also analyzed the relationship between the latency and the DEP in mMIMO systems. Then, Zeng *et al.* extended the results in [15] to mMIMO systems with shadow fading, demonstrating that mMIMO can provide URLLC services for multiple devices even suffering from severe shadow fading [16]. The optimal secure performance was obtained by optimizing the channel blocklength and transmission bits per packet in [17]. The pilot power and payload power was jointly optimized to maximize the weighted sum rate of multiple devices in a single cell [18]. The authors of [19] considered a more general scenario of multiple cells with imperfect channel state information (CSI) and pilot contamination, and showed that the pilot contamination had a significant impact on the reliability of URLLC services. Although it has been shown that mMIMO can provide URLLC services for multiple devices, it may be unable to provide guaranteed URLLC services to all devices in the cell due to blockage issue and severe inter-cell interference. Therefore, a novel network architecture should be developed to support URLLC services.

By geographically deploying distributed APs, cell-free mMIMO (CF mMIMO) can provide uniform services for all devices [20]–[22]. The performance improvements of CF MIMO systems over the centralized mMIMO systems have been shown when using maximum ratio transmission (MRT) precoding scheme [23] and zero forcing (ZF) precoding scheme [24], respectively. Considering that the previous precoding schemes may no longer be applicable for CF mMIMO, the authors of [25] proposed four precoding schemes, namely, full-pilot zero-forcing (FZF), local partial zero-forcing precoding, local protective partial zero-forcing, and local regularized zero-forcing. The energy efficiency of CF mMIMO was analyzed in [26]. The aforementioned works in [20]–[25] assumed that the APs can acquire the CSI of all devices, which is theoretically possible but impractical. To address this issues, a user-centric approach was proposed to reduce the implementation complexity [27]. The coverage probability with various

densities of APs was analyzed in [28]. To tackle the blockage issue, the performance of the reconfigurable intelligent surface-aided CF mMIMO system was analyzed in [29]. However, all works were based on the assumption of infinite channel blocklength, which is not suitable for short packet transmission.

Due to the appealing advantages of CF mMIMO, it has great potential to provide URLLC services for multiple devices simultaneously in a large coverage area. Essentially, there was a significant improvement in terms of the network's availability over the centralized mMIMO [30]. The power allocation based on FCBL for maximizing the minimal data rate and maximizing the energy efficiency was considered in [31], where each AP was equipped with a single antenna. However, channel hardening can only be achieved by deploying ultra-high density of single-antenna APs [32], which is theoretically possible but practically unrealistic due to the expensive hardware. In this paper, we investigate the deployment of multiple-antenna APs and optimal AP selection under the short packet regime. Then, we aim to maximize the weighted sum rate based on FCBL while considering the minimal requirements of DEP and data rate, by optimizing the pilot power and the transmission power. The main contributions of this paper are summarized as follows.

- 1) By using the user-centric approach, we derive the lower bounds (LBs) of the achievable downlink data rate with imperfect CSI for the MRT, FZF, and local zero-forcing (LZF) precoding schemes when using FCBL.
- 2) The weighted sum rate is maximized by jointly optimizing the pilot power and the transmission power while considering the minimal requirements of DEP and data rate. To solve this NP-hard problem, we first transform the DEP and data rate requirements into the required SINR, and then reducing the number of variables by proving that the globally optimal solution of pilot power can be derived in closed form. Furthermore, by introducing the approximations, the problem can be simplified into a series of subproblems, which can be transformed into a geometric programming (GP) problem by using log-function method and successive convex approximation (SCA) [33], [34]. Finally, an iterative algorithm is proposed to solve this problem for three linear precoding schemes.
- 3) Simulation results demonstrate the rapid convergence speed of our proposed algorithms, and also validate the effectiveness of our method over the existing algorithm. Besides, by accessing various APs, the optimal AP selection strategy based on short packet transmission is provided. More importantly, the CF mMIMO system has a remarkable performance improvement over the centralized mMIMO system.

The remainder of this paper is organized as follows. In Section II, the system model is provided, and then the LB data rate expression under FCBL based on statistical CSI is derived for the MRT, FZF, and LZF precoding schemes, respectively. In Section III, the power allocation is optimized to maximize the ergodic sum data rate. Then, simulation results are presented in Section IV. Finally, the conclusions are drawn in Section V.

Notation: The superscripts $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ stand for the conjugate, transpose, and conjugate-transpose, respectively. The Euclidean norm and the expectation operator are denoted by $\|\cdot\|$ and $\mathbb{E}\{\cdot\}$, respectively. $z \sim \mathcal{CN}(0, 1)$ denotes a circularly symmetric complex Gaussian random variable (RV) z with zero mean and unit variance, and $\mathbf{z} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ means an N -dimensional complex vector, each element of which is independent and follows the distribution of $\mathcal{CN}(0, 1)$. Finally, $\mathbf{A} \in \mathbb{C}^{M \times N}$ means that \mathbf{A} is a complex matrix with M rows and N columns.

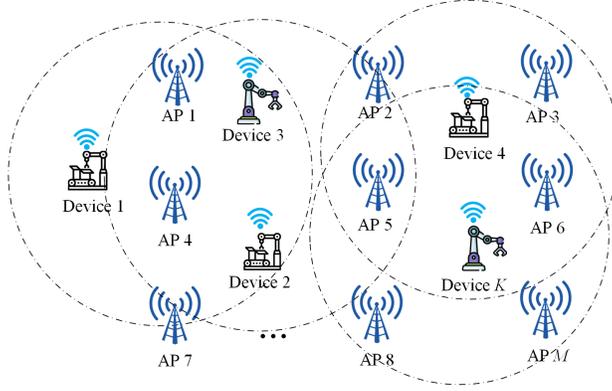


Fig. 1: Smart factory scenario where CF mMIMO serves multiple devices.

II. SYSTEM MODEL AND SPECTRAL EFFICIENCY

A. System Model

We consider a CF mMIMO-enabled smart factory where M APs equipped with N antennas jointly serve all K single-antenna devices, as illustrated in Fig. 1. The channel vector $\mathbf{g}_{m,k} \in \mathbb{C}^{N \times 1}$ between the m th AP and the k th device is modeled as

$$\mathbf{g}_{m,k} = \sqrt{\beta_{m,k}} \mathbf{h}_{m,k}, \quad (1)$$

where $\beta_{m,k}$ is the large-scale fading and $\mathbf{h}_{m,k} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_N)$ denotes a normal distribution with zero mean and variance of \mathbf{I}_N .

B. Uplink Training

It is assumed that each AP needs to estimate the CSI from all the devices based on time division duplex (TDD) protocol within the limited channel blocklength $L = B \times T_B$, where B is the bandwidth and T_B is the transmission duration. In order to distinguish the channels from different devices, K devices are allocated with orthogonal pilot sequences. Then, the m th AP estimates the channel matrix based on the received pilot signal $\mathbf{Y}_m^p \in \mathbb{C}^{N \times K}$, which is given by

$$\mathbf{Y}_m^p = \sum_{k=1}^K \mathbf{g}_{m,k} \sqrt{K p_k^p} \mathbf{q}_k^H + \mathbf{N}_m^p, \quad (2)$$

where p_k^p is the pilot power of the k th device, $\mathbf{q}_k \in \mathbb{C}^{K \times 1}$ is the k th device's pilot sequence, and $\mathbf{N}_m^p \in \mathbb{C}^{N \times K}$ is the additive Gaussian noise matrix at the m th AP, each element of which is independent and follows the distribution of $\mathcal{CN}(0, 1)$. By multiplying (2) with orthogonal pilot \mathbf{q}_k , we have

$$\hat{\mathbf{y}}_{m,k}^p = \frac{1}{\sqrt{K p_k^p}} \mathbf{Y}_m^p \mathbf{q}_k = \mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p, \quad (3)$$

where $\mathbf{n}_{m,k}^p = \frac{1}{\sqrt{K p_k^p}} \mathbf{N}_m^p \mathbf{q}_k$. Based on (3), the estimated channel vector $\hat{\mathbf{g}}_{m,k}$ by using minimum mean-square error (MMSE) is

$$\hat{\mathbf{g}}_{m,k} = \frac{K p_k^p \beta_{m,k}}{K p_k^p \beta_{m,k} + 1} \hat{\mathbf{y}}_{m,k}^p, \quad (4)$$

which follows the distribution of $\mathcal{CN}(\mathbf{0}, \lambda_{m,k} \mathbf{I}_N)$ with $\lambda_{m,k}$ given by

$$\lambda_{m,k} = \frac{K p_k^p (\beta_{m,k})^2}{K p_k^p \beta_{m,k} + 1}. \quad (5)$$

Then, let us denote $\tilde{\mathbf{g}}_{m,k} = \mathbf{g}_{m,k} - \hat{\mathbf{g}}_{m,k}$ as the channel estimation error, which is independent of $\hat{\mathbf{g}}_{m,k}$ and follows the distribution of $\mathcal{CN}(\mathbf{0}, (\beta_{m,k} - \lambda_{m,k}) \mathbf{I}_N)$.

C. Downlink Transmission

For downlink transmission, to reduce the computational complexity, the user-centric approach is adopted, e.g., each device is served by a subset of APs or each AP serves a subset of devices. Denote \mathcal{M}_k as the set of APs that serve the k th device and \mathcal{U}_m as the set of devices that are served by the m th AP, respectively. The transmitted signal from the m th AP is denoted as

$$\mathbf{x}_m = \sum_{k' \in \mathcal{U}_m} \sqrt{p_{m,k'}^d} \mathbf{a}_{m,k'}^* s_{k'}, \quad (6)$$

where $p_{m,k'}^d$ is the transmission power, $\mathbf{a}_{m,k'}$ is the precoding vector, and $s_{k'}$ is the data symbol to the k' th device.

The received signal at the k th device is

$$\begin{aligned} y_k^d &= \sum_{m=1}^M \sum_{k' \in \mathcal{U}_m} \sqrt{p_{m,k'}^d} \mathbf{g}_{m,k}^T \mathbf{a}_{m,k'}^* s_{k'} + n_k \\ &= \sum_{k'=1}^K \sum_{m \in \mathcal{M}_{k'}} (\mathbf{g}_{m,k})^T \mathbf{a}_{m,k'}^* \sqrt{p_{m,k'}^d} s_{k'} + n_k, \end{aligned} \quad (7)$$

where n_k is the noise with the distribution of $\mathcal{CN}(0, 1)$. Besides, since there are no downlink pilots, we assume that the k th device treats the mean of the effect channel gain as the true channel for signal detection [35]. Then, the received signal at the k th device can be rewritten as

$$\begin{aligned} y_k^d &= \mathbb{E} \left\{ \underbrace{\sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T \mathbf{a}_{m,k}^* \sqrt{p_{m,k}^d}}_{\text{DS}_k} \right\} s_k \\ &\quad + \underbrace{\left\{ \sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T \mathbf{a}_{m,k}^* \sqrt{p_{m,k}^d} - \text{DS}_k \right\}}_{\text{LS}_k} s_k, \\ &\quad + \underbrace{\sum_{k' \neq k} \sum_{m \in \mathcal{M}_{k'}} (\mathbf{g}_{m,k})^T \mathbf{a}_{m,k'}^* \sqrt{p_{m,k'}^d} s_{k'}}_{\text{UI}_{k,k'}} + \underbrace{n_k}_{\text{N}_k}, \end{aligned} \quad (8)$$

where DS_k is the desired signal, LS_k is the leaked signal, $\text{UI}_{k,k'}$ represents the interference due to the k' th device, and N_k is the noise term. The SINR of the k th device is given by

$$\gamma_k = \frac{|\text{DS}_k|^2}{|\text{LS}_k|^2 + \sum_{k' \neq k} |\text{UI}_{k,k'}|^2 + |\text{N}_k|^2}. \quad (9)$$

For the precoding vector $\mathbf{a}_{m,k}$, we consider the following three linear precoding schemes [25], [36]

$$\mathbf{a}_{m,k} = \begin{cases} \frac{\hat{\mathbf{G}}_m \mathbf{e}_k}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{G}}_m \mathbf{e}_k\|^2\}}}, & \text{MRT} \\ \frac{\hat{\mathbf{G}}_m (\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m)^{-1} \mathbf{e}_k}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{G}}_m [\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m]^{-1} \mathbf{e}_k\|^2\}}}, & \text{FZF} \\ \frac{\hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} (\mathbf{E}_{\mathcal{U}_m}^H \hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m})^{-1} \boldsymbol{\xi}_{m,k}}{\sqrt{\mathbb{E}\{\|\hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} (\mathbf{E}_{\mathcal{U}_m}^H \hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m})^{-1} \boldsymbol{\xi}_{m,k}\|^2\}}}, & \text{LZF} \end{cases} \quad (10)$$

where $\mathbb{E}\{\cdot\}$ denotes the expectation operator, $\hat{\mathbf{G}}_m = [\hat{\mathbf{g}}_{m,1}, \hat{\mathbf{g}}_{m,2}, \dots, \hat{\mathbf{g}}_{m,K}]$ is the estimated channel matrix between all the devices and the m th AP, and \mathbf{e}_k represents the k th column of unit matrix \mathbf{I}_K . For the LZF precoding scheme, $\hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} = [\hat{\mathbf{g}}_{m,d_1}, \hat{\mathbf{g}}_{m,d_2}, \dots, \hat{\mathbf{g}}_{m,d_{|\mathcal{U}_m|}}] \in \mathbb{C}^{N \times |\mathcal{U}_m|}$ is a matrix collecting the channels of serving devices in \mathcal{U}_m , where $\mathcal{U}_m = \{d_1, d_2, \dots, d_{|\mathcal{U}_m|}\}$ is the set of devices served by the m th AP and $\mathbf{E}_{\mathcal{U}_m}$ is $[\mathbf{e}_{d_1}, \mathbf{e}_{d_2}, \dots, \mathbf{e}_{d_{|\mathcal{U}_m|}}] \in \mathbb{C}^{K \times |\mathcal{U}_m|}$. For ease of exposition, let $\mathcal{U}_m^{\text{index}} = \{1, 2, \dots, |\mathcal{U}_m|\}$ be the set comprised of the index of \mathcal{U}_m . Given user k , we can find an index $j \in \mathcal{U}_m^{\text{index}}$ where $d_j = k$. Then, we have $\boldsymbol{\xi}_{m,k} = [\mathbf{I}_{|\mathcal{U}_m|}]_{(:,j)}$.

As can be seen from (10), for the FZF precoding scheme, the m th AP needs to estimate all devices' channels, and thus it can suppress the interference of all devices by sacrificing spatial degrees of freedom. In contrast, the m th AP using the MRT and the LZF precoding methods only needs to know the serving devices' CSI, which reduces the implementation complexity. Besides, the system based on the LZF precoder can only suppresses the interference causing by serving devices, which strikes a balance between the available spatial degrees of freedom and the interference suppression.

D. Achievable Data Rate under Finite Blocklength

Based on Shannon's coding theorem, the Shannon capacity is defined as the maximum coding rate that there exists an encoder/decoder pair that can enable the DEP to approach zero when the channel blocklength is infinity [37]. However, in short packet transmission, the DEP has a non-negligible impact on the data rate. In [4], the authors derived the approximate achievable data rate for the k th device under FCBL, which is given by

$$R_k \approx (1 - \eta) \log_2(1 + \gamma_k) - \sqrt{\frac{(1 - \eta) V_k(\gamma_k) Q^{-1}(\varepsilon_k)}{L \ln 2}}, \quad (11)$$

where $\eta = K/L$, γ_k is the k th device's SINR, ε_k is DEP, V_k is the channel dispersion with $V_k(\gamma_k) = 1 - (1 + \gamma_k)^{-2}$, and $Q^{-1}(\varepsilon_k)$ is the inverse function of $Q(\varepsilon_k) = \frac{1}{\sqrt{2\pi}} \int_{\varepsilon_k}^{\infty} e^{-t^2/2} dt$ of the k th device.

The ergodic data rate of the k th device under FCBL is given by

$$\begin{aligned} \bar{R}_k &\approx \mathbb{E} \left\{ \frac{1 - \eta}{\ln 2} \left[\ln(1 + \gamma_k) - \frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1 - \eta)}} \sqrt{\frac{\frac{2}{\gamma_k} + 1}{\left(\frac{1}{\gamma_k} + 1\right)^2}} \right] \right\}, \\ &\triangleq \frac{1 - \eta}{\ln 2} \mathbb{E} \left\{ f_k \left(\frac{1}{\gamma_k} \right) \right\}, \end{aligned} \quad (12)$$

where $f_k(x) = \ln(1 + \frac{1}{x}) - \frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1 - \eta)}} \sqrt{\frac{2x + 1}{(1 + x)^2}}$ is a function based on the k th device's DEP requirements, and the expectation is taken over the small-scale fading channel. As can be seen from (12), the closed-form expression of the ergodic data rate is challenging to derive, and thus we cannot allocate the power based on the exact expression

of (12). To address this issue, we aim to derive the LB of the ergodic data rate which is more convenient for resource allocation.

Assuming that the data rate R_k of any device is no smaller than 0, we have the following inequality

$$\frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}} \leq \frac{\left(\frac{1}{\gamma_k} + 1\right) \ln(1 + \gamma_k)}{\sqrt{\frac{2}{\gamma_k} + 1}} \triangleq g\left(\frac{1}{\gamma_k}\right), \quad (13)$$

where $g(x)$ is equal to $\frac{(1+x)\ln(1+\frac{1}{x})}{\sqrt{2x+1}}$. We can readily check that the first-order derivative of $g(x)$ is smaller than 0, and thus $g(x)$ is a monotonically decreasing function. Besides, the feasible region of $f_k(x)$ is $0 \leq x \leq g^{-1}\left(\frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}}\right)$. As a result, we have the following lemma.

Lemma 1: Function $f_k(x)$ is a decreasing and convex function when $0 < x \leq g^{-1}\left(\frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}}\right)$.

Proof: Please refer to Appendix B in [38]. ■

By using Jensen's inequality and Lemma 1, we have

$$\bar{R}_k \geq \hat{R}_k \triangleq \frac{1-\eta}{\ln 2} f_k(1/\hat{\gamma}_k), \quad (14)$$

where \hat{R}_k is the LB data rate of the k th device, and $\hat{\gamma}_k$ is $\hat{\gamma}_k = \frac{1}{\mathbb{E}(1/\gamma_k)}$.

To obtain the closed-form expression of \hat{R}_k , the k th device's SINRs based on the MRT, FZF, and LZP precoding schemes should be derived. Specifically, we have the following results.

Theorem 1: The ergodic achievable data rate for the k th device using the MRT precoding scheme under FCBL can be lower bounded by

$$\hat{R}_k^{\text{MRT}} \triangleq \frac{1-\eta}{\ln 2} f_k\left(\frac{1}{\hat{\gamma}_k^{\text{MRT}}}\right), \quad (15)$$

where $\hat{\gamma}_k^{\text{MRT}}$ is denoted as

$$\hat{\gamma}_k^{\text{MRT}} = \frac{\left(\sum_{m \in \mathcal{M}_k} \sqrt{N p_{m,k}^d \lambda_{m,k}}\right)^2}{\sum_{k'=1}^K \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \beta_{m,k} + 1}. \quad (16)$$

Proof: Please refer to Appendix A. ■

Theorem 2: Using the FZF precoding scheme, the k th device's ergodic data rate is lower bounded by

$$\hat{R}_k^{\text{FZF}} \triangleq \frac{1-\eta}{\ln 2} f_k\left(\frac{1}{\hat{\gamma}_k^{\text{FZF}}}\right), \quad (17)$$

where $\hat{\gamma}_k^{\text{FZF}}$ is denoted as

$$\hat{\gamma}_k^{\text{FZF}} = \frac{\left(\sum_{m \in \mathcal{M}_k} \sqrt{(N-K) p_{m,k}^d \lambda_{m,k}}\right)^2}{\sum_{k'=1}^K \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d (\beta_{m,k} - \lambda_{m,k}) + 1}, \quad (18)$$

where the number of antennas N should be larger than the number of devices K .

Proof: Please refer to Appendix B. ■

Theorem 3: The k th device's ergodic data rate based on the LZP precoding scheme is lower bounded by

$$\hat{R}_k^{\text{LZF}} \triangleq \frac{1-\eta}{\ln 2} f_k\left(\frac{1}{\hat{\gamma}_k^{\text{LZF}}}\right), \quad (19)$$

where $\hat{\gamma}_k^{\text{LZF}}$ is given by

$$\hat{\gamma}_k^{\text{LZF}} = \frac{\left(\sum_{m \in \mathcal{M}_k} \sqrt{(N - \tau_m) p_{m,k}^d \lambda_{m,k}} \right)^2}{\sum_{k'=1}^K \left[\sum_{m \in \{\mathcal{M}_{k'} \cap \mathcal{M}_k\}} p_{m,k'}^d (\beta_{m,k} - \lambda_{m,k}) + \sum_{m \in \{\mathcal{M}_{k'} \setminus \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}\}} p_{m,k'}^d \beta_{m,k} \right] + 1}. \quad (20)$$

In (20), τ_m means the number of devices served by the m th AP and its value is given by $\tau_m = |\mathcal{U}_m|$. Here, the number of antennas N should be larger than τ_m .

Proof: Please refer to Appendix C. ■

From the expressions of SINRs in (18) and (20), the FZF precoding scheme is a special case of the LZFP precoder, i.e., the device is served by all APs. Besides, we also note that τ_m is always no larger than the number of devices K , as \mathcal{U}_m is a subset of devices. Therefore, by choosing the optimal set of APs, it is reasonable for the system to adopt the LZFP precoding scheme to support more devices than that based on the FZF precoder.

III. POWER ALLOCATION

In this section, we aim to optimize the power allocation to maximize the weighted sum rate.

A. Problem Formulation

We assume that all the devices have the same bandwidth B , and we aim to maximize the weighted sum rate with limited energy constraints and the minimal data rate requirement. Mathematically, the optimization problem can be formulated as

$$\max_{\{p_k^p\}, \{p_{m,k}^d\}} \sum_{k=1}^K w_k \hat{R}_k \quad (21a)$$

$$\text{s.t.} \quad \hat{R}_k \geq R_k^{\text{req}}, \forall k, \quad (21b)$$

$$p_k^p \leq P_k^{\text{max},p}, \forall k \quad (21c)$$

$$\sum_{k \in \mathcal{U}_m} p_{m,k}^d \leq P_m^d, \forall m, \quad (21d)$$

where \hat{R}_k denotes the LB data rate based on the abovementioned three precoding schemes, R_k^{req} is the k th device's data rate requirement, w_k is the weight of the k th device, $P_k^{\text{max},p}$ is the maximal power of the k th device, P_m^d is the m th AP's maximal transmission power. Specifically, constraint (21b) means the k th device's minimal data rate requirements, constraint (21c) and constraint (21d) mean that the uplink training power of each device and the total transmission power of each AP are limited.

For the power allocation based on infinite blocklength in [39], [40], the problem can be converted into a convex problem by introducing slack variables, which can be readily solved by the bisection search algorithm. However, maximizing the weighted sum rate is an NP-hard problem, which cannot be readily solved. Besides, it is more challenging to solve the weighted sum rate problem under imperfect CSI and FCBL. Therefore, we first simplify the problem, and then propose an efficient algorithm for solving the problem with polynomial-time complexity.

Using Lemma 1, the minimal data rate requirement in (21b) can be transformed into the k th device's requirement of SINR, denoted as

$$\hat{\gamma}_k \geq \frac{1}{f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1-\eta}\right)}, \quad (22)$$

where $\hat{\gamma}_k$ represents the k th device's SINR using the abovementioned precoding schemes. Besides, we find the globally optimal solution for pilot power based on the following lemma.

Lemma 2: $f_k(\frac{1}{\hat{\gamma}_k})$ is a monotonically increasing function of pilot power p_k^p when $0 < \frac{1}{\hat{\gamma}_k} \leq g^{-1}\left(\frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}}\right)$.

Proof: Please refer to Appendix D. ■

By using (22) and substituting $p_k^p = P_k^{\text{max},p}$ into the SINR's expression, Problem (21) can be simplified as

$$\max_{\{p_{m,k}^d\}} \sum_{k=1}^K w_k \hat{R}_k \quad (23a)$$

$$\text{s.t. } \hat{\gamma}_k \geq \frac{1}{f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1-\eta}\right)}, \forall k, \quad (23b)$$

$$\sum_{k \in \mathcal{U}_m} p_{m,k}^d \leq P_m, \forall m. \quad (23c)$$

Then, by introducing slack variables χ_k , Problem (21) can be equivalently transformed into the following optimization problem

$$\max_{\{p_{m,k}^d\}, \{\chi_k\}} \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} [\ln(1+\chi_k) - \alpha_k G(\chi_k)] \quad (24a)$$

$$\text{s.t. } \hat{\gamma}_k \geq \chi_k, \forall k, \quad (24b)$$

$$\chi_k \geq \frac{1}{f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1-\eta}\right)}, \forall k, \quad (24c)$$

$$(23c), \quad (24d)$$

where $G(\chi_k)$ is defined as $G(\chi_k) \triangleq \sqrt{\frac{\frac{2}{\chi_k} + 1}{\left(\frac{1}{\chi_k} + 1\right)^2}}$, and α_k is $\alpha_k = \frac{Q^{-1}(\varepsilon_k)}{\sqrt{L(1-\eta)}}$.

To further simplify the objective function in (24a), the following lemmas are introduced.

Lemma 3: For any given $\hat{x} \geq 0$, function $\ln(1+x)$ can be lower bounded by

$$\ln(1+x) \geq \rho \ln x + \delta, \quad (25)$$

where ρ and δ are expressed as

$$\rho = \frac{\hat{x}}{1+\hat{x}}, \delta = \ln(1+\hat{x}) - \frac{\hat{x}}{1+\hat{x}} \ln(\hat{x}). \quad (26)$$

Proof: Please refer to Appendix E. ■

Lemma 4: For any given $\hat{x} \geq \frac{\sqrt{17}-3}{4}$, function $G(x)$ always satisfies the following inequality

$$G(x) \leq \hat{\rho} \ln(x) + \hat{\delta}, \quad (27)$$

where $\hat{\rho}$ and $\hat{\delta}$ are given by

$$\hat{\rho} = \frac{\hat{x}}{\sqrt{\hat{x}^2 + 2\hat{x}}} - \frac{\hat{x}\sqrt{\hat{x}^2 + 2\hat{x}}}{(1+\hat{x})^2}, \quad (28)$$

and

$$\hat{\delta} = \sqrt{1 - \frac{1}{(1 + \hat{x})^2}} - \hat{\rho} \ln(\hat{x}). \quad (29)$$

Proof: Please refer to Appendix D in [18]. ■

By using Lemma 3 and Lemma 4, the weighted data rate can be approximated in an iterative manner, which is detailed as follows

$$\begin{aligned} & w_k \frac{(1 - \eta)}{\ln 2} [\ln(1 + \gamma_k) - \alpha_k G(\gamma_k)] \\ & \geq w_k \frac{(1 - \eta)}{\ln 2} \left[\rho_k^{(i)} \ln(\gamma_k) + \delta_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \ln(\gamma_k) - \alpha_k \hat{\delta}_k^{(i)} \right], \end{aligned} \quad (30)$$

where $\rho_k^{(i)}$, $\delta_k^{(i)}$, $\hat{\rho}_k^{(i)}$, and $\hat{\delta}_k^{(i)}$ are obtained based on (26), (28), and (29) by using $\hat{x} = \gamma_k^{(i)}$ in the i th iteration. As a result, the weighted sum rate in (24a) can be lower bounded by

$$\begin{aligned} & \sum_{k=1}^K w_k \frac{(1 - \eta)}{\ln 2} [\ln(1 + \chi_k) - \alpha_k G(\chi_k)] \\ & \geq \sum_{k=1}^K w_k \frac{(1 - \eta)}{\ln 2} \left[\ln(\chi_k)^{\left[\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \right]} + \delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)} \right], \end{aligned} \quad (31)$$

where the equality holds only when $\chi_k = \chi_k^{(i)}$.

Next, we focus on the term consisting of variable χ_k in (31), and solve the following subproblem in the i th iteration

$$\max_{\{p_{m,k}^d\}, \{\chi_k\}} \prod_{k=1}^K \chi_k^{\hat{w}_k^{(i)}} \quad (32a)$$

$$\text{s.t.} \quad (24b), (24c), (23c), \quad (32b)$$

where $\hat{w}_k^{(i)}$ is equal to $\hat{w}_k^{(i)} = w_k \frac{(1 - \eta)}{\ln 2} (\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)})$.

Obviously, the problem in (32) is not a GP problem as constraint (24b) is not a monomial function [41]. To tackle this issue, considering the different expressions for various precoding schemes, we denote the numerator and the denominator of SINR $\hat{\gamma}_k$ as $(\theta_k)^2$ and ϖ_k , respectively, and then introduce a general theorem to approximate θ_k based on abovementioned three precoding schemes as a monomial function.

Theorem 4: For any given $\hat{p}_{m,k}^d > 0$, θ_k is lower bounded by

$$\begin{aligned} \theta_k &= \sum_{m \in \mathcal{M}_k} \sqrt{(N - t_m) p_{m,k}^d \hat{\lambda}_{m,k}} \\ &\geq c_k \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^d \hat{\lambda}_{m,k} \right]^{a_{m,k}}, \end{aligned} \quad (33)$$

where $\hat{\lambda}_{m,k}$ is equal to $\hat{\lambda}_{m,k} = \frac{K P_k^{\max, p} (\beta_{m,k})^2}{K P_k^{\max, p} \beta_{m,k} + 1}$, t_m is a constant that depends on the different precoding schemes, $a_{m,k}$ and c_k are the coefficients. Specifically, t_m is give by

$$t_m = \begin{cases} 0, & \text{MRT} \\ K, & \text{FZF} \\ \tau_m, & \text{LZF} \end{cases}. \quad (34)$$

The coefficients $a_{m,k}$ and c_k are given by

$$a_{m,k} = \frac{\sqrt{(N - t_m) \hat{p}_{m,k}^d \hat{\lambda}_{m,k}}}{2\hat{\theta}_k}, \quad (35)$$

and

$$c_k = \frac{\hat{\theta}_k}{\prod_{m \in \mathcal{M}_k} \left[(N - t_m) \hat{p}_{m,k}^d \hat{\lambda}_{m,k} \right]^{a_{m,k}}}, \quad (36)$$

where $\hat{\theta}_k$ is obtained by using $p_{m,k}^d = \hat{p}_{m,k}^d$. Besides, it is obvious that the inequality in (33) holds with equality when $p_{m,k}^d = \hat{p}_{m,k}^d$.

Proof: Please refer to Appendix F. ■

By using Theorem 4, similar to the objection function in (24a), we can approximate the numerator $(\theta_k)^2$ in an iterative manner. Specifically, $c_k^{(i)}$ and $a_{m,k}^{(i)}$ are obtained based on (35) and (36) by using $\hat{p}_{m,k}^d = p_{m,k}^{d,(i)}$, and θ_k can be lower bounded by

$$\theta_k \geq c_k^{(i)} \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^d \hat{\lambda}_{m,k} \right]^{a_{m,k}^{(i)}}. \quad (37)$$

Based on the abovementioned simplifications and approximations, the problem is transformed into the following GP problem

$$\max_{\{p_{m,k}^d\}, \{\chi_k\}} \prod_{k=1}^K \chi_k \hat{w}_k^{(i)} \quad (38a)$$

$$\begin{aligned} \text{s.t.} \quad & \left(c_k^{(i)} \right)^2 \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^d \hat{\lambda}_{m,k} \right]^{2a_{m,k}^{(i)}} \\ & \geq \chi_k \varpi_k, \forall k, \end{aligned} \quad (38b)$$

$$(24c), (23c). \quad (38c)$$

To run the iterative algorithm, it is necessary to find a feasible initial solution. To deal with this issue, we construct an alternative optimization problem by introducing an auxiliary variable φ , which is given by

$$\max_{\varphi, \{p_{m,k}^d\}} \varphi \quad (39a)$$

$$\begin{aligned} \text{s.t.} \quad & \left(c_k^{(i)} \right)^2 \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^d \hat{\lambda}_{m,k} \right]^{2a_{m,k}^{(i)}} \\ & \geq \frac{\varphi}{f_k^{-1} \left(\frac{R_k^{\text{req}} \ln 2}{1 - \eta} \right)} \varpi_k, \end{aligned} \quad (39b)$$

$$(23c). \quad (39c)$$

Obviously, Problem (39) is also a GP problem, and is always feasible. Besides, Problem (38) is feasible only when φ is no smaller than 1. Furthermore, we define an error tolerance ξ to guarantee that the transmission power converges to the optimal solutions. Based on the abovementioned discussions, Algorithm 1 is provided to maximize the weighted sum rate.

Algorithm 1 Iterative Algorithm for Solving Maximum Weighted Sum Rate

- 1: Initialize iteration number $i = 1$, and error tolerance $\zeta = 0.01$;
 - 2: Initialize the pilot power $\{p_k^p = P_k^{\max,p}, \forall k\}$, calculate transmission power $\{p_{m,k}^{d,(1)}, \forall m, k\}$ by solving Problem (39), obtain SINR $\{\chi_k^{(1)}, \forall k\}$ and the weighted sum rate in (21a) denoted as $\text{Obj}^{(1)}$. Set $\text{Obj}^{(0)} = \text{Obj}^{(1)}\zeta$;
 - 3: **while** $(\text{Obj}^{(i)} - \text{Obj}^{(i-1)})/\text{Obj}^{(i-1)} \geq \zeta$ **do**
 - 4: Update $\{\hat{w}_k^{(i)}, c_k^{(i)}, a_{m,k}^{(i)}, \forall m, k\}$;
 - 5: Update $i = i + 1$, solve Problem (38) by using the CVX package to obtain $\{p_{m,k}^{d,(i)}, \forall m, k\}$, calculate SINR $\{\chi_k^{(i)}, \forall k\}$ and then obtain the weighted sum rate, denoted as $\text{Obj}^{(i)}$;
 - 6: **end while**
-

B. Algorithm Analysis

1) *Feasibility Analysis*: For Algorithm 1, we need to check whether constraint (38b) in the i th iteration holds or not in the $(i + 1)$ th iteration as only constraint (38b) is approximated in an iterative manner. Constraint (38b) in the i th iteration is given by

$$\left(c_k^{(i-1)}\right)^2 \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^{d,(i)} \hat{\lambda}_{m,k}\right]^{2a_{m,k}^{(i-1)}} \geq \chi_k^{(i)} \varpi_k^{(i)}, \quad (40)$$

where $\{\chi_k^{(i)}, p_{m,k}^{d,(i)}, \forall m, k\}$ is the optimal solution in the i th iteration, and $\varpi_k^{(i)}$ is obtained by using $p_{m,k}^d = p_{m,k}^{d,(i)}$.

Using Theorem 4 and (37), we have

$$\begin{aligned} c_k^{(i)} \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^{d,(i)} \hat{\lambda}_{m,k}\right]^{a_{m,k}^{(i)}} &= \theta_k^{(i)} \\ &\geq c_k^{(i-1)} \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^{d,(i)} \hat{\lambda}_{m,k}\right]^{a_{m,k}^{(i-1)}}. \end{aligned} \quad (41)$$

Then, by combining (40) with (41), we have

$$\left(c_k^{(i)}\right)^2 \prod_{m \in \mathcal{M}_k} \left[(N - t_m) p_{m,k}^{d,(i)} \hat{\lambda}_{m,k}\right]^{2a_{m,k}^{(i)}} \geq \chi_k^{(i)} \varpi_k^{(i)}. \quad (42)$$

Obviously, the solution is also feasible in the $(i + 1)$ th iteration.

2) *Convergence Analysis*: We prove that our algorithm can converge to a locally optimal solution. Denote $\text{Obj}^{(i)}$ as the weighted sum rate in the i th iteration. Since the solution in the i th iteration is also feasible in the $(i + 1)$ th iteration, we have

$$\begin{aligned} &\sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(\chi_k^{(i+1)} \right)^{\left[\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \right]} + \delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)} \right] \\ &\geq \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(\chi_k^{(i)} \right)^{\left[\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \right]} + \delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)} \right] \\ &= \text{Obj}^{(i)}, \end{aligned} \quad (43)$$

where $\{\chi_k^{(i+1)}, \forall k\}$ is the optimal solution to Problem (38) in the $(i + 1)$ th iteration.

Substituting $\chi_k = \chi_k^{(i+1)}$ into the inequality in (31), we have

$$\begin{aligned}
& \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(1 + \chi_k^{(i+1)} \right) - \alpha_k G \left(\chi_k^{(i+1)} \right) \right] \\
& \geq \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \times \\
& \quad \left[\ln \left(\chi_k^{(i+1)} \right)^{\left[\rho_k^{(i+1)} - \alpha_k \hat{\rho}_k^{(i+1)} \right]} + \delta_k^{(i+1)} - \alpha_k \hat{\delta}_k^{(i+1)} \right] \\
& \geq \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(\chi_k^{(i+1)} \right)^{\left[\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \right]} + \delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)} \right].
\end{aligned} \tag{44}$$

Then, by combining (43) with (44), we have

$$\begin{aligned}
& \text{Obj}^{(i+1)} \\
& = \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(1 + \chi_k^{(i+1)} \right) - \alpha_k G \left(\chi_k^{(i+1)} \right) \right] \\
& \geq \sum_{k=1}^K w_k \frac{(1-\eta)}{\ln 2} \left[\ln \left(\chi_k^{(i+1)} \right)^{\left[\rho_k^{(i)} - \alpha_k \hat{\rho}_k^{(i)} \right]} + \delta_k^{(i)} - \alpha_k \hat{\delta}_k^{(i)} \right] \\
& \geq \text{Obj}^{(i)}.
\end{aligned} \tag{45}$$

Therefore, the convergence of Algorithm 1 is verified. Besides, we can prove that Algorithm 1 can converge to the Karush-Kuhn-Tucker (KKT) point of Problem (21) for the abovementioned precoding schemes by using the similar proof as in Appendix B in [42].

3) *Complexity Analysis*: The complexity of Algorithm 1 depends on the number of iterations and complexity of each iteration. Specifically, the main complexity of each iteration in Algorithm 1 lies in solving Problem (38) which includes $(M+1)K$ variables and $(2K+M)$ constraints. Based on [41], the computational complexity of this algorithm is on the order of $\mathcal{O}(N_{iter} \times \max\{[(M+1)K]^3, (2K+M)[(M+1)K]^2, N_{cost}\})$, where N_{iter} is the number of iterations and N_{cost} is the computational complexity of calculating the first-order and second-order derivatives of the objective function and constraint functions of Problem (38) [34]. Furthermore, our simulation results demonstrate that Algorithm 1 can converge to the locally optimal solution with fewer iterations.

IV. SIMULATION RESULTS

The performance of the proposed algorithms are numerically evaluated and discussed in this section. We first introduce the simulation setup and the related simulation parameters.

A. Simulation Scenario

The smart factory is assumed to be located in a $D \times D$ square. In contrast to the wraparound deployment in [43], [44], we uniformly deploy M APs at constellation points to provide uniform service for the devices. The

large-scale fading coefficient model is adopted [23], which is given by

$$\text{PL}_{m,k} = \begin{cases} L_{\text{loss}} + 35\log_{10}(d_{m,k}), & d_{m,k} > d_1, \\ L_{\text{loss}} + 15\log_{10}(d_1) + 20\log_{10}(d_0), & d_{m,k} \leq d_0, \\ L_{\text{loss}} + 15\log_{10}(d_1) + 20\log_{10}(d_{m,k}), & \text{other,} \end{cases} \quad (46)$$

where $d_{m,k}$ (km) is the distance between the m th AP and the k th device, and L_{loss} (dB) is a constant factor that depends on the carrier frequency f (MHz), the heights of the APs h_{AP} (m) and devices h_u (m). Specifically, L_{loss} is given by

$$\begin{aligned} L_{\text{loss}} = & 46.3 + 33.9\log_{10}(f) - 13.82\log_{10}(h_{\text{AP}}) \\ & - (1.1\log_{10}(f) - 0.7)h_u + (1.56\log_{10}(f) - 0.8). \end{aligned} \quad (47)$$

Besides, for the small-scale fading, it is generally modeled as Rayleigh fading with zero mean and unit variance. The corresponding normalized pilot power p_k^p and transmission power $p_{m,k}^d$ can be computed through dividing these powers by the noise power, which is given by

$$P_n = B \times k_B \times T_0 \times 10^{\frac{N_{\text{dB}}}{10}} \text{ (W)}, \quad (48)$$

where $k_B = 1.381 \times 10^{-23}$ (Joule per Kelvin) is the Boltzmann constant, and $T_0 = 290$ (Kelvin) is the noise temperature. The weights for all the devices are randomly generated within $[0,1]$. Unless otherwise specified, the simulation parameters are similar to those in [25], [45] and summarized in Table I. More importantly, we fix the total number of antennas in this smart factory to investigate the deployment of APs. In other words, if each AP is equipped with more antennas, this area will deploy less APs.

As mentioned before, the k th device is served by the set of APs \mathcal{M}_k . Specifically, it is assumed that the large-scale fading factors are known at the m th AP, and then the large-scale fading factors $\{\beta_{1,k}, \beta_{2,k}, \dots, \beta_{M,k}\}$ are sorted in a descending order. Finally, the large-scale fading factors are selected in turn until satisfying the following condition

$$\frac{\sum_{m \in \mathcal{M}_k} \beta_{m,k}}{\sum_{m=1}^M \beta_{m,k}} \geq T_h, \quad (49)$$

where T_h is the threshold. For the set of devices served by the m th AP, by checking whether the m th AP belongs to the set of \mathcal{M}_k , $k = 1, 2, \dots, K$, we can obtain \mathcal{U}_m .

B. Properties of the Proposed Algorithm

In this subsection, we first check the gap between the LB data rate and the ergodic data rate, illustrate the convergence behavior of the proposed algorithm, and then investigate the impact of threshold on the system performance.

1) *Tightness*: The simulation results are obtained through the Monte-Carlo simulation by averaging over 10^4 random channel generations with $T_h = 0.9$ and $p_{m,k}^d = 0.1 \text{ W}$, $\forall m, k$. As can be seen from Fig. 2, the derived LB data rate is close to the ergodic rate for any system parameters, which confirms that the LB data rate is suitable and reasonable for power allocation.

TABLE I: Simulation Parameters

Parameters Setting	Value
Carrier frequency (f)	2.1 GHz
Bandwidth (B)	10 MHz
Transmission duration (T_B)	0.05 ms
Blocklength ($L = BT_B$)	500
Height of APs (h_{AP})	15 m
Height of devices (h_u)	1.6 m
Noise figure (N_{dB})	9 dB
Number of devices (K)	10
Required data rate (R_{req})	0.5 bit/s/Hz
Decoding error probability ε_k	10^{-7}
Size of square (D)	1000 m
Pilot power $P_k^{\max}, \forall k$	100 mW
d_0	10 m
d_1	50 m

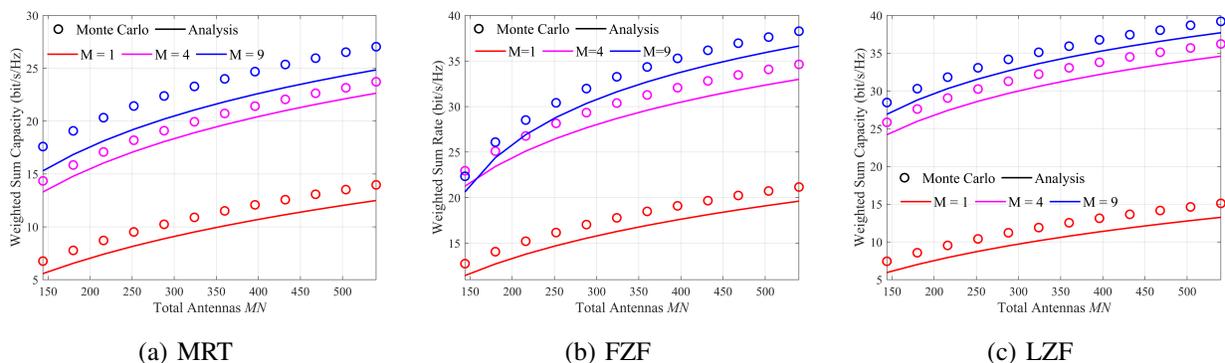


Fig. 2: Weighted Sum Rate V.S. The Number of Total Antennas under various numbers of APs.

2) *Convergence*: We investigate the convergence behavior of the proposed algorithm with $MN = 144$ in Fig. 3. For given any transmission power P_m and threshold T_h , the system performance for three precoding schemes can converge to the locally optimal solution within only 2 or 3 iterations, which demonstrates the rapid convergence of the proposed algorithm.

3) *Threshold*: The performance of the proposed algorithm is obtained by averaging 100 random devices' locations and the system performance is set to zero if any devices cannot satisfy the data requirements. Fig. 4 shows the system performance versus different thresholds with $P_m = 1 W, \forall m$. Obviously, it is observed that the optimal value of T_h is 1 for the MRT and FZF schemes and 0.95 for the LZF precoder. This is due to the fact that selecting more APs to provide service for devices will consume the degrees of freedom for the LZF precoding scheme, leading to performance degradation. Here, we set $T_h = 0.95$ for all the following simulations, to achieve a good tradeoff between system performance and computational complexity.

C. Effect of pilot power

In this subsection, we investigate how pilot power affects the system performance. Fig. 5 depicts the average weighted sum rate versus the pilot power with $MN = 144$ and $P_m = 1 W, \forall m$, by averaging 100 random devices' locations. As can be seen from Fig. 5, the average weighted sum rate increases with the pilot power for any cases,

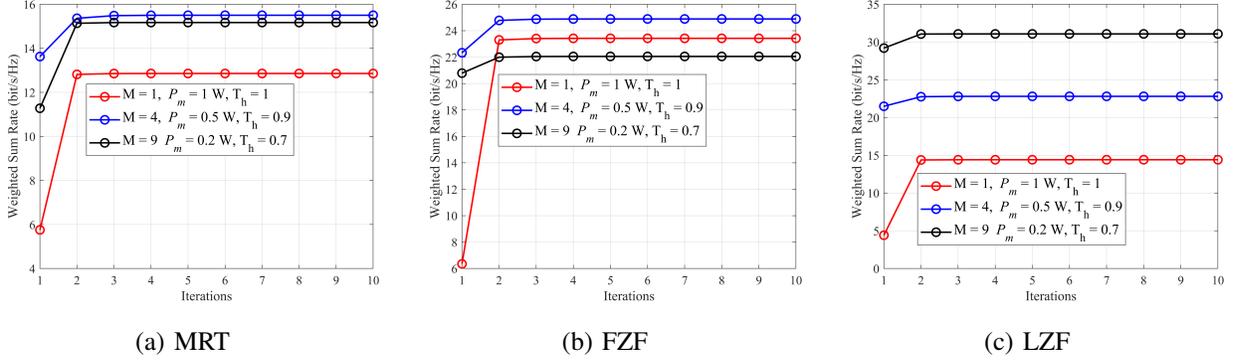


Fig. 3: Convergence of proposed algorithm for different precoding schemes.

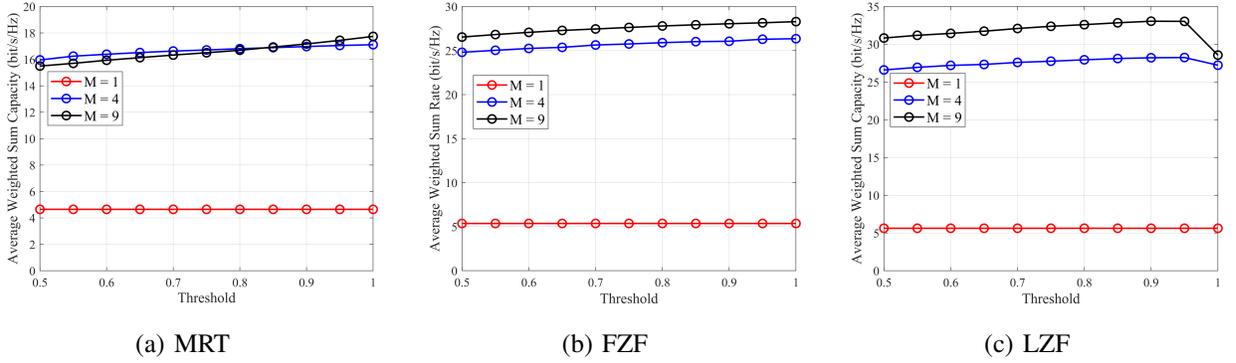


Fig. 4: Performance of the proposed algorithm versus threshold for different precoding schemes.

which demonstrates that more accurate channel estimation is beneficial for enhancing the system performance. More importantly, we find an interesting phenomenon that the CF mMIMO (e.g., $M \geq 4$) significantly outperforms the centralized mMIMO system (e.g., $M = 1$) when the pilot power is low. This is attributed to the fact that the devices are closer to the APs in CF mMIMO systems than in centralized mMIMO systems, hence less pilot power is required to satisfy the requirements of DEP and data rate.

D. Effect of The Number of APs

To fully explore the deployment of APs so as to maximize the system performance with limited antennas, we evaluate the average weighted sum rate versus various numbers of APs with $MN = 144$ and $P_m = 0.2W, \forall m$ in Fig. 6. For the MRT scheme, the average weighted sum rate initially increases with the number of APs, and then it tends to be stable at 16 bit/s/Hz. This is due to the fact that each device relying on the MRT precoding scheme becomes interference limited and tends to be stable. However, for the FZF precoding scheme, the system performance will decrease when the number of APs is large, as deploying more APs causes the reduction in degrees of freedom. In contrast, the system performance using the LZF precoding scheme increases with the number of APs. This is because the LZF precoding scheme strikes a balance between interference suppression and available degrees of freedom, thereby supporting more devices.

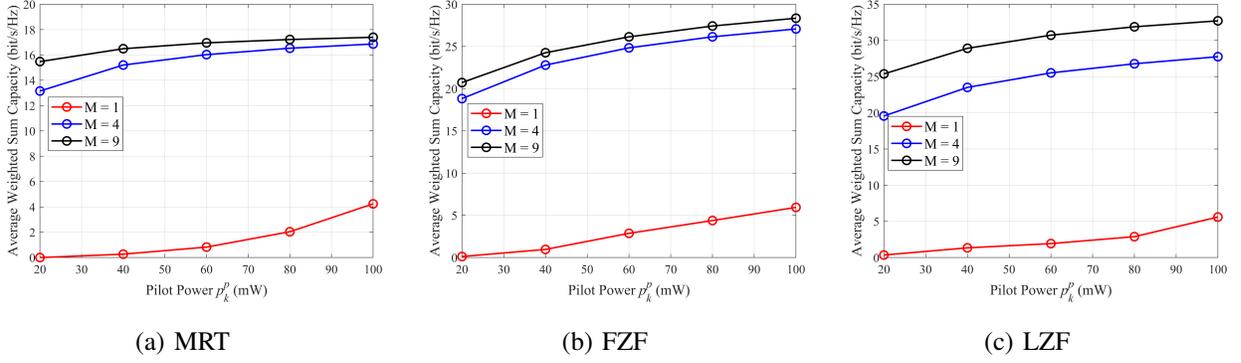


Fig. 5: Performance of the algorithm versus pilot power for different precoding schemes.

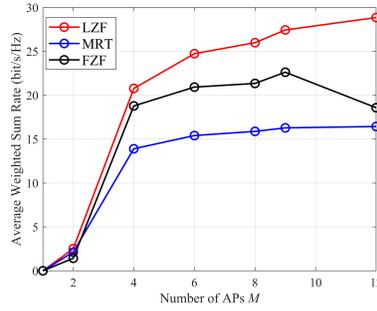


Fig. 6: Performance of proposed algorithm V.S. Number of APs.

E. Effect of The Number of Devices

To support more devices, a round robin-based scheduler is adopted for the case of $K \geq N$ or $K \geq |\mathcal{U}_m|$. Specifically, the APs would first transmit signals to the K_1 devices, and then serve the remaining $(K - K_1)$ devices in the next time interval. By averaging over 100 random generations, we investigate the relationship between the number of devices and the system performance with $MN = 144$, $K_1 = \frac{K}{2}$, and $P_m = 0.2 W$, $\forall m$. To show the effectiveness of our proposed method, the results of the Shannon capacity, the algorithm in [11], and power allocation in [25] are presented. Furthermore, if any devices violate the requirements, the data rate is set to zero. Obviously, the Shannon capacity is the ideal performance, and the benchmark one in [11] has an unpredictable trend because it does not consider the penalty due to short packet transmissions. The performance relying on the power allocation of [25] can approach that of the proposed method, owing to enhanced path gain. In contrast, the proposed algorithm can approach the upper bound in CF mMIMO systems, which demonstrates the effectiveness of our algorithm. More importantly, the weighted sum rate in the centralized mMIMO is almost zero owing to the failure to meet the requirements, while there is a significant performance improvement in CF mMIMO systems. This is due to the fact that the centralized mMIMO can only support those devices that are close to the APs, instead of all devices, leading to zero data rate. Furthermore, the average weighted sum rate of $M = 9$ APs relying on the FZF precoding increases when $K \leq 10$ and then declines when $10 \leq K \leq 14$. Thereafter, APs based on the FZF

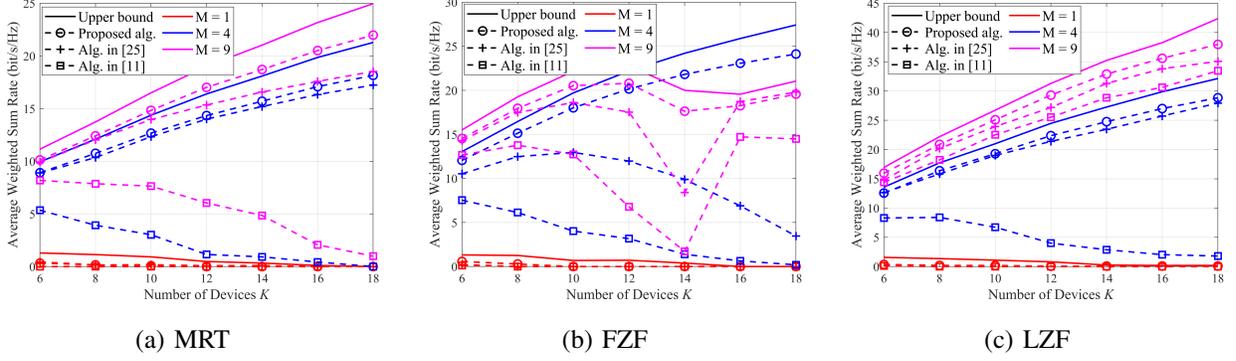


Fig. 7: Performance of proposed algorithm V.S. Number of devices for different precoding schemes.

precoding scheme can support extra devices with enhanced URLLC services by implementing the scheduler, which motivates us to enforce the appropriate scheduler when the number of served devices approaches that of equipped antennas per AP.

V. CONCLUSION

In this paper, the resource allocation for a CF mMIMO-enabled URLLC downlink system was treated. We first derived the closed-form LB data rates with imperfect CSI based on MRT, FZF, and LZF precoding, and maximized the weighted sum rate based on the derived LB data rate. Then, by deriving the globally optimal pilot power and using SCA, the non-convex problem was transformed into a series of subproblems, which can be solved in an iterative manner by our proposed algorithm. Simulation results demonstrated the rapid convergence speed of our algorithm and the optimal AP selection strategy based on the short packet transmission. Furthermore, the quality of URLLC services will benefit by deploying more APs, except for the FZF precoding scheme. More importantly, the power allocation strategies under the short packet regime can significantly enhance the system performance over the existing algorithms.

Regarding CF mMIMO systems, it is impractical to assign orthogonal pilot sequences to multiple devices under the FCBL. Therefore, investigating the pilot allocation scheme and analyzing the impact of sharing pilot sequences would be left for our future work. Furthermore, since it is unrealistic to assume an idealized fronthaul link between the CPU and the APs, the limited fronthaul will be studied in the future.

APPENDIX A
PROOF OF THEOREM 1

Before proving this theorem, we need to calculate the precoding vector for the MRT case. The normalized precoding vector is given by

$$\begin{aligned} \mathbf{a}_{m,k}^{\text{MRT}} &= \frac{\alpha_{m,k} \left(\mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p \right)}{\sqrt{\mathbb{E} \left\| \alpha_{m,k} \left(\mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p \right) \right\|^2}} = \frac{\mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p}{\sqrt{N \left(\beta_{m,k} + \frac{1}{K p_k^p} \right)}} \\ &= \frac{\sqrt{\lambda_{m,k}}}{\beta_{m,k} \sqrt{N}} \left(\mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p \right), \end{aligned} \quad (50)$$

where $\alpha_{m,k}$ is $\alpha_{m,k} = \frac{K p_k^p \beta_{m,k}}{K p_k^p \beta_{m,k} + 1}$.

Then, we need to derive the expressions of $|\text{DS}_k|^2$, $\mathbb{E} \left(|\text{LS}_k|^2 \right)$, $\mathbb{E} \left(|\text{UI}_{k,k'}|^2 \right)$ and $\mathbb{E} \left(|\text{N}_k|^2 \right)$, respectively. We first compute DS_k . Since $\hat{\mathbf{g}}_{m,k}$ and $\tilde{\mathbf{g}}_{m,k}$ are independent, we have

$$\begin{aligned} &|\text{DS}_k|^2 \\ &= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} \sqrt{p_{m,k}^d} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{MRT}})^* \right\} \right|^2 \\ &= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} \sqrt{p_{m,k}^d} (\mathbf{g}_{m,k})^T \frac{\sqrt{\lambda_{m,k}}}{\beta_{m,k} \sqrt{N}} \left(\mathbf{g}_{m,k} + \mathbf{n}_{m,k}^p \right)^* \right\} \right|^2 \\ &= \left| \sum_{m \in \mathcal{M}_k} \sqrt{N p_{m,k}^d \lambda_{m,k}} \right|^2. \end{aligned} \quad (51)$$

The term $\mathbb{E} \left(|\text{LS}_k|^2 \right)$ is given by

$$\begin{aligned} &\mathbb{E} \left\{ |\text{LS}_k|^2 \right\} \\ &= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} \sqrt{p_{m,k}^d} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{MRT}})^* - \text{DS}_k \right|^2 \right\} \\ &+ \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} \frac{\sqrt{p_{m,k}^d \lambda_{m,k}}}{\beta_{m,k} \sqrt{N}} (\mathbf{g}_{m,k})^T (\mathbf{n}_{m,k}^p)^* \right|^2 \right\} \\ &- \left(\sum_{m \in \mathcal{M}_k} \sqrt{N \lambda_{m,k} p_{m,k}^d} \right)^2 \\ &= \sum_{m \in \mathcal{M}_k} p_{m,k}^d \beta_{m,k}. \end{aligned} \quad (52)$$

Then, $\mathbb{E}(|\text{UI}_{k,k'}|^2)$ can be calculated as

$$\begin{aligned} & \mathbb{E}(|\text{UI}_{k,k'}|^2) \\ &= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} \frac{\sqrt{p_{m,k'}^d \gamma_{m,k'}}}{\beta_{m,k'} \sqrt{N}} (\mathbf{g}_{m,k})^T (\mathbf{g}_{m,k'})^* \right|^2 \right\} \\ &+ \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} \frac{\sqrt{p_{m,k'}^d \gamma_{m,k'}}}{\beta_{m,k'} \sqrt{N}} (\mathbf{g}_{m,k})^T (\mathbf{n}_{m,k'}^p)^* \right|^2 \right\}. \end{aligned} \quad (53)$$

For each term in (53), we have

$$\begin{aligned} & \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} \frac{\sqrt{p_{m,k'}^d \gamma_{m,k'}}}{\beta_{m,k'} \sqrt{N}} (\mathbf{g}_{m,k})^T (\mathbf{g}_{m,k'})^* \right|^2 \right\} \\ &= \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \gamma_{m,k'} \frac{\beta_{m,k}}{\beta_{m,k'}} \end{aligned} \quad (54)$$

and

$$\begin{aligned} & \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} \frac{\sqrt{p_{m,k'}^d \gamma_{m,k'}}}{\beta_{m,k'} \sqrt{N}} (\mathbf{g}_{m,k})^T (\mathbf{n}_{m,k'}^p)^* \right|^2 \right\} \\ &= \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \gamma_{m,k'} \frac{1}{K p_{k'}^p} \frac{\beta_{m,k}}{(\beta_{m,k'})^2}. \end{aligned} \quad (55)$$

By combining (54) with (55), we have

$$\begin{aligned} & \mathbb{E}(|\text{UI}_{k,k'}|^2) \\ &= \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \gamma_{m,k'} \frac{\beta_{m,k}}{\beta_{m,k'}} + \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \gamma_{m,k'} \frac{\beta_{m,k}}{(\beta_{m,k'})^2} \frac{1}{K p_{k'}^p} \\ &= \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d \beta_{m,k}. \end{aligned} \quad (56)$$

Finally, we compute $\mathbb{E}(|N_k|^2)$, which is written as

$$\mathbb{E}\{|n_k|^2\} = 1. \quad (57)$$

Substituting (51), (52), (56), and (57) into (9), we obtain $\hat{\gamma}_k^{\text{MRT}}$ in (16).

APPENDIX B
PROOF OF THEOREM 2

Before proving this theorem, we need to provide the precoding vector. By using the identity [46], the normalized coefficient can be derived as

$$\begin{aligned}
& \mathbb{E} \left\{ \left\| \hat{\mathbf{G}}_m \left[\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \right]^{-1} \mathbf{e}_k \right\|^2 \right\} \\
&= \mathbb{E} \left\{ (\mathbf{e}_k)^H \left[\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \right]^{-1} \left(\hat{\mathbf{G}}_m \right)^H \hat{\mathbf{G}}_m \left[\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \right]^{-1} \mathbf{e}_k \right\} \\
&= \mathbb{E} \left\{ (\mathbf{e}_k)^H \left[\hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \right]^{-1} \mathbf{e}_k \right\} \\
&= \frac{1}{(N - K) \lambda_{m,k}}.
\end{aligned} \tag{58}$$

Then, $|\text{DS}_k|$ can be derived as

$$\begin{aligned}
& |\text{DS}_k|^2 \\
&= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{FZF}})^* \sqrt{p_{m,k}^d} \right\} \right|^2 \\
&= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} (\hat{\mathbf{g}}_{m,k} + \tilde{\mathbf{g}}_{m,k})^T (\mathbf{a}_{m,k}^{\text{FZF}})^* \sqrt{p_{m,k}^d} \right\} \right|^2 \\
&= \left(\sum_{m \in \mathcal{M}_k} \sqrt{(N - K) p_{m,k}^d \lambda_{m,k}} \right)^2.
\end{aligned} \tag{59}$$

Next, the leakage power can be formulated as

$$\begin{aligned}
& \mathbb{E} \left\{ |\text{LS}_k|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{FZF}})^* \sqrt{p_{m,k}^d} - \text{DS}_k \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} (\tilde{\mathbf{g}}_{m,k})^T (\mathbf{a}_{m,k}^{\text{FZF}})^* \sqrt{p_{m,k}^d} \right|^2 \right\} \\
&= \sum_{m \in \mathcal{M}_k} p_{m,k}^d (\beta_{m,k} - \lambda_{m,k}).
\end{aligned} \tag{60}$$

The term $\mathbb{E} \left(|\text{UI}_{k,k'}|^2 \right)$ can be expressed as

$$\begin{aligned}
& \mathbb{E} \left\{ |\text{UI}_{k,k'}|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{FZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} (\tilde{\mathbf{g}}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{FZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \sum_{m \in \mathcal{M}_{k'}} p_{m,k'}^d (\beta_{m,k} - \lambda_{m,k}).
\end{aligned} \tag{61}$$

Finally, we complete the proof by substituting the expressions of (59), (60), (61), and $\mathbb{E} \left\{ |n_k|^2 \right\} = 1$ into the SINR expression.

APPENDIX C PROOF OF THEOREM 3

The normalized coefficient can be derived as

$$\begin{aligned}
& \mathbb{E} \left\{ \left\| \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} \left(\mathbf{E}_{\mathcal{U}_m}^H \hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} \right)^{-1} \boldsymbol{\xi}_{m,k} \right\|^2 \right\} \\
&= \mathbb{E} \left\{ (\boldsymbol{\xi}_{m,k})^H \left(\mathbf{E}_{\mathcal{U}_m}^H \hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} \right)^{-1} \boldsymbol{\xi}_{m,k} \right\} \\
&= \frac{1}{(N - \tau_m) \lambda_{m,k}},
\end{aligned} \tag{62}$$

where τ_m is defined in (34).

Then, the desired signal $|\text{DS}_k|^2$ can be given by

$$\begin{aligned}
& |\text{DS}_k|^2 \\
&= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{LZF}})^* \sqrt{p_{m,k}^d} \right\} \right|^2 \\
&= \left| \mathbb{E} \left\{ \sum_{m \in \mathcal{M}_k} \sqrt{(N - \tau_m) \lambda_{m,k} p_{m,k}^d} (\hat{\mathbf{g}}_{m,k})^T \left(\hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} \left(\mathbf{E}_{\mathcal{U}_m}^H \hat{\mathbf{G}}_m^H \hat{\mathbf{G}}_m \mathbf{E}_{\mathcal{U}_m} \right)^{-1} \boldsymbol{\xi}_{m,k} \right)^* \right\} \right|^2 \\
&= \left(\sum_{m \in \mathcal{M}_k} \sqrt{(N - \tau_m) p_{m,k}^d \lambda_{m,k}} \right)^2.
\end{aligned} \tag{63}$$

Next, similar to the FZF case, the leakage power for the AP using the LZF precoding scheme can be formulated

as

$$\begin{aligned}
& \mathbb{E} \left\{ |\text{LS}_k|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k}^{\text{LZF}})^* \sqrt{p_{m,k}^d} - \text{DS}_k \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_k} (\tilde{\mathbf{g}}_{m,k})^T (\mathbf{a}_{m,k}^{\text{LZF}})^* \sqrt{p_{m,k}^d} \right|^2 \right\} \\
&= \sum_{m \in \mathcal{M}_k} p_{m,k}^d (\beta_{m,k} - \lambda_{m,k}).
\end{aligned} \tag{64}$$

The term of the devices' interference is different from that of the FZF scheme, as the interference from other devices may not be suppressed. The term $\mathbb{E} \left(|\text{UI}_{k,k'}|^2 \right)$ can be given by

$$\begin{aligned}
\mathbb{E} \left\{ |\text{UI}_{k,k'}|^2 \right\} &= \mathbb{E} \left\{ \left| \sum_{m \in \mathcal{M}_{k'}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} + \sum_{m \in \{\mathcal{M}_{k'} \setminus \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 + \left| \sum_{m \in \{\mathcal{M}_{k'} \setminus \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\}.
\end{aligned} \tag{65}$$

As can be seen from (65), the devices' interference consists of two terms. In specific, the first term means that the vector $\hat{\mathbf{g}}_{m,k}$ is chosen by the selection matrix $\mathbf{E}_{\mathcal{U}_m}$ and the second term means that not chosen by matrix $\mathbf{E}_{\mathcal{U}_m}$. Obviously, the interference of the first term can be suppressed as $\hat{\mathbf{g}}_{m,k} \mathbf{a}_{m,k'}^{\text{LZF}}$ is equal to zero, while the second term's interference cannot be suppressed.

Then, the first and the second terms of $\mathbb{E} \left(|\text{UI}_{k,k'}|^2 \right)$ are given by

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \sum_{m \in \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \mathbb{E} \left\{ \left| \sum_{m \in \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}} (\tilde{\mathbf{g}}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \sum_{m \in \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}} p_{m,k'}^d (\beta_{m,k} - \lambda_{m,k}),
\end{aligned} \tag{66}$$

and

$$\begin{aligned}
& \mathbb{E} \left\{ \left| \sum_{m \in \{\mathcal{M}_{k'} \setminus \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}\}} (\mathbf{g}_{m,k})^T (\mathbf{a}_{m,k'}^{\text{LZF}})^* \sqrt{p_{m,k'}^d} \right|^2 \right\} \\
&= \sum_{m \in \{\mathcal{M}_{k'} \setminus \{\mathcal{M}_k \cap \mathcal{M}_{k'}\}\}} p_{m,k'}^d \beta_{m,k}.
\end{aligned} \tag{67}$$

Finally, the k th device's SINR using the LZF precoding scheme is obtained by substituting the expressions of

(63), (64), (65), and $\mathbb{E}\{|n_k|^2\} = 1$ into (9).

APPENDIX D

PROOF OF LEMMA 2

We first derive the first-order derivative of $f_k(\frac{1}{\hat{\gamma}_k})$, denoted as

$$\frac{df_k\left(\frac{1}{\hat{\gamma}_k}\right)}{dp_k^p} = \frac{-f_k'\left(\frac{1}{\hat{\gamma}_k}\right) d\hat{\gamma}_k}{(\hat{\gamma}_k)^2 dp_k^p}. \quad (68)$$

As can be seen, the sign of the first-order derivative depends on the sign of $\frac{d\hat{\gamma}_k}{dp_k^p}$. Due to the different SINR expressions of the three precoding schemes, we define a general expression $\hat{\gamma}_k = Y(\lambda_{m,k})$. Then, it is readily to prove that $\hat{\gamma}_k$ monotonically increases with $\lambda_{m,k}$, and the first-order derivative of $\lambda_{m,k}$ is $\frac{K(\beta_{m,k})^2}{(Kp_k^p\beta_{m,k}+1)^2} > 0$. Therefore, the function $f_k(\frac{1}{\hat{\gamma}_k})$ monotonically increases with pilot power p_k^p , and the data rate can be maximized when $p_k^p = P_k^{\max,p}$.

APPENDIX E

PROOF OF LEMMA 3

The inequality in (25) can be readily proved by substituting the expressions of ρ and δ into (25). Then, we define $J(x) = \ln(1+x) - \rho \ln x - \delta$, the first-order derivative is given by

$$\frac{dJ(x)}{dx} = \frac{x - \rho(1+x)}{(1+x)x} = \frac{x(1+\hat{x}) - \hat{x}(1+x)}{(1+\hat{x})(1+x)x}. \quad (69)$$

Since both x and \hat{x} are positive values, the sign of $\frac{dJ(x)}{dx}$ only depends on the numerator. Let us define $H(x) = x(1+\hat{x}) - \hat{x}(1+x)$, and then the first-order derivative of $H(x)$ is given by $H'(x) = 1$, which means $H(x)$ monotonically increases. Consequently, since $H(\hat{x})$ is equal to zero, we have $H(x) \geq 0$ when $x \geq \hat{x}$ and $H(x) \leq 0$ when $x \leq \hat{x}$, which indicates that $J(x)$ is an increasing function when $x \geq \hat{x}$ and a decreasing function when $x \leq \hat{x}$. As a result, we complete the proof by showing that $J(x)$ is always larger than $J(\hat{x}) = 0$.

APPENDIX F

PROOF OF THEOREM 4

By taking the logarithm operator for the left hand side of (33), we have

$$\ln(\theta_k) = \ln\left(\sum_{m \in \mathcal{M}_k} \sqrt{(N-t_m)p_{m,k}^d \hat{\lambda}_{m,k}}\right) \triangleq F(\mathbf{x}), \quad (70)$$

where \mathbf{x} is given by $\mathbf{x} = [\ln[(N-t_1)p_{1,k}^d \hat{\lambda}_{1,k}], \dots, \ln[(N-t_m)p_{m,k}^d \hat{\lambda}_{m,k}]]^T$, $m \in \mathcal{M}_k$.

The first-order partial derivative of $F(\mathbf{x})$ is given by

$$\frac{\partial F(\mathbf{x})}{\partial x_{j,k}} = \frac{\sqrt{e^{x_{j,k}}}}{2 \sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}}} = \frac{\sqrt{(N-t_j)p_{j,k}^d \hat{\lambda}_{j,k}}}{2\theta_k}, \quad (71)$$

where e is the exponent. The second-order partial derivatives of $F(\mathbf{x})$ are given by

$$\frac{\partial^2 F(\mathbf{x})}{\partial(x_{j,k})^2} = \frac{\sqrt{e^{x_{j,k}}} \left(\sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}} \right) - (\sqrt{e^{x_{j,k}}})^2}{4 \left(\sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}} \right)^2}, \quad (72)$$

and

$$\frac{\partial^2 F(\mathbf{x})}{\partial x_{j,k} \partial x_{i,k}} = \frac{-\sqrt{e^{x_{j,k}}} e^{x_{i,k}}}{4 \left(\sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}} \right)^2}. \quad (73)$$

Then, we define $\mathbf{z}_k = [\sqrt{e^{x_{1,k}}}, \sqrt{e^{x_{2,k}}}, \dots, \sqrt{e^{x_{|\mathcal{M}_k|,k}}}]^T$, $m \in \mathcal{M}_k$, and thus the Hessian matrix of $F(\mathbf{x})$ can be given by

$$\begin{aligned} \mathbf{H} &= \frac{1}{4(\mathbf{1}^T \mathbf{z}_k)^2} \left\{ \left(\sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}} \right) \begin{bmatrix} \sqrt{e^{x_{1,k}}} & \dots & 0 \\ 0 & \sqrt{e^{x_{2,k}}} & 0 \\ 0 & \dots & \sqrt{e^{x_{|\mathcal{M}_k|,k}}} \end{bmatrix} - \mathbf{z}_k (\mathbf{z}_k)^T \right\} \\ &= \frac{1}{4(\mathbf{1}^T \mathbf{z}_k)^2} \underbrace{\left\{ \mathbf{1}^T \mathbf{z}_k \text{diag} \{ \mathbf{z}_k \} - \mathbf{z}_k (\mathbf{z}_k)^T \right\}}_{\Xi}, \end{aligned} \quad (74)$$

In (74), $\mathbf{1}$ is a vector of $[1, 1, \dots, 1]^T$, $|\mathcal{M}_k|$ means the cardinality of the set \mathcal{M}_k . For any given $\mathbf{v} = [v_1, \dots, v_{|\mathcal{M}_k|}]^T \in \mathbb{R}^{|\mathcal{M}_k|}$, by using the Cauchy-Schwartz inequality, we have the inequality that is given by

$$\begin{aligned} \mathbf{v}^T \Xi \mathbf{v} &= \mathbf{1}^T \mathbf{z}_k \mathbf{v}^T \text{diag} \{ \mathbf{z}_k \} \mathbf{v} - \mathbf{v}^T \mathbf{z}_k (\mathbf{z}_k)^T \mathbf{v} \\ &= \left(\sum_{m \in \mathcal{M}_k} \sqrt{e^{x_{m,k}}} \right) \left(\sum_{m \in \mathcal{M}_k} v_m \sqrt{e^{x_{m,k}}} v_m \right) - \left(\sum_{m \in \mathcal{M}_k} v_m \sqrt{e^{x_{m,k}}} \right)^2 \\ &= \left(\sum_{m \in \mathcal{M}_k} \left(\sqrt{\sqrt{e^{x_{m,k}}}} \right)^2 \right) \left(\sum_{m \in \mathcal{M}_k} \left(v_m \sqrt{\sqrt{e^{x_{m,k}}}} \right)^2 \right) - \left(\sum_{m \in \mathcal{M}_k} v_m \sqrt{e^{x_{m,k}}} \right)^2 \geq 0. \end{aligned} \quad (75)$$

Therefore, we prove $\ln(\theta_k)$ is a convex function of \mathbf{x} . Then, by using Jensen's inequality, we have

$$F(x) \geq \sum_{m \in \mathcal{M}_k} a_{m,k} x_{m,k} + \ln(c_k), \quad (76)$$

where $a_{m,k}$ and c_k are given in (35) and (36), respectively.

Finally, we complete the proof by taking the exponential operation for both sides of (76) and using $x_{m,k} = \ln(p_{m,k}^d)$.

REFERENCES

- [1] A. A. Kumar S., K. Ovsthus, and L. M. Kristensen., "An industrial perspective on wireless sensor networks — a survey of requirements, protocols, and challenges," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1391–1412, 3rd Quart. 2014.
- [2] P. Popovski, Č. Stefanović, J. J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A.-S. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.
- [3] H. Ren, C. Pan, Y. Deng, M. El Kashlan, and A. Nallanathan, "Joint power and blocklength optimization for URLLC in a factory automation scenario," *IEEE Trans. Wireless Commun.*, vol. 19, no. 3, pp. 1786–1801, Dec. 2020.

- [4] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [5] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sept. 2016.
- [6] C. She, C. Yang, and T. Q. S. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, Jun. 2017.
- [7] C. Pan, H. Ren, Y. Deng, M. ElKashlan, and A. Nallanathan, "Joint blocklength and location optimization for URLLC-enabled UAV relay systems," *IEEE Commun. Lett.*, vol. 23, no. 3, pp. 498–501, Mar. 2019.
- [8] A. A. Nasir, "Min-max decoding-error probability-based resource allocation for a URLLC system," *IEEE Commun. Lett.*, vol. 24, no. 12, pp. 2864–2867, Dec. 2020.
- [9] X. Xie, X. Ou, H. Lu, and Q. Huang, "Joint uplink and downlink resource allocation in NOMA for end-to-end URLLC services," *IEEE Commun. Lett.*, vol. 25, no. 12, pp. 3942–3946, Dec. 2021.
- [10] J. Wan, B. Chen, M. Imran, F. Tao, D. Li, C. Liu, and S. Ahmad, "Toward dynamic resources management for IoT-based manufacturing," *IEEE Commun. Mag.*, vol. 56, no. 2, pp. 52–59, Feb. 2018.
- [11] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink URLLC-OFDMA systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, 2019, pp. 1–6.
- [12] V. K. Huang, Z. Pang, C.-J. A. Chen, and K. F. Tsang, "New trends in the practical deployment of industrial wireless: From noncritical to critical use cases," *IEEE Ind. Electron. Mag.*, vol. 12, no. 2, pp. 50–58, Jun. 2018.
- [13] H. Q. Ngo, E. G. Larsson, and T. L. Marzetta, "Energy and spectral efficiency of very large multiuser MIMO systems," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1436–1449, Apr. 2013.
- [14] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.
- [15] J. Zeng, T. Lv, R. P. Liu, X. Su, N. C. Beaulieu, and Y. J. Guo, "Linear minimum error probability detection for massive MU-MIMO with imperfect CSI in URLLC," *IEEE Trans. Veh. Technol.*, vol. 68, no. 11, pp. 11 384–11 388, Nov. 2019.
- [16] J. Zeng, T. Lv, R. P. Liu, X. Su, Y. J. Guo, and N. C. Beaulieu, "Enabling ultrareliable and low-latency communications under shadow fading by massive MU-MIMO," *IEEE Int. Things J.*, vol. 7, no. 1, pp. 234–246, Jan. 2020.
- [17] T. Yu, X. Sun, and Y. Cai, "Secure short-packet transmission in uplink massive MU-MIMO IoT networks," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Oct. 2020, pp. 50–55.
- [18] H. Ren, C. Pan, Y. Deng, M. ElKashlan, and A. Nallanathan, "Joint pilot and payload power allocation for massive-MIMO-enabled URLLC IIoT networks," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 5, pp. 816–830, May 2020.
- [19] J. Östman, A. Lancho, G. Durisi, and L. Sanguinetti, "URLLC with massive MIMO: Analysis and design at finite blocklength," *IEEE Trans. Wireless Commun.*, vol. 20, no. 10, pp. 6387–6401, Oct. 2021.
- [20] T. M. Hoang, H. Q. Ngo, T. Q. Duong, H. D. Tuan, and A. Marshall, "Cell-free massive MIMO networks: Optimal power control against active eavesdropping," *IEEE Trans. Commun.*, vol. 66, no. 10, pp. 4724–4737, Oct. 2018.
- [21] H. Q. Ngo, L.-N. Tran, T. Q. Duong, M. Matthaiou, and E. G. Larsson, "On the total energy efficiency of cell-free massive MIMO," *IEEE Trans. Green Commun. Netw.*, vol. 2, no. 1, pp. 25–39, Mar. 2018.
- [22] M. Bashar, K. Cumanan, A. G. Burr, M. Debbah, and H. Q. Ngo, "On the uplink Max-Min SINR of cell-free massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 18, no. 4, pp. 2021–2036, Apr. 2019.
- [23] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [24] P. Liu, K. Luo, D. Chen, and T. Jiang, "Spectral efficiency analysis of cell-free massive MIMO systems with zero-forcing detector," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 795–807, Feb. 2020.
- [25] G. Interdonato, M. Karlsson, E. Björnson, and E. G. Larsson, "Local partial zero-forcing precoding for cell-free massive MIMO," *IEEE Trans. Wireless Commun.*, vol. 19, no. 7, pp. 4758–4774, Jul. 2020.
- [26] A. Papazafeiropoulos, H. Q. Ngo, P. Kourtessis, S. Chatzinotas, and J. M. Senior, "Towards optimal energy efficiency in cell-free massive MIMO systems," *IEEE Trans. Green Commun. and Netw.*, vol. 5, no. 2, pp. 816–831, Jun. 2021.
- [27] E. Björnson and L. Sanguinetti, "Scalable cell-free massive MIMO systems," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4247–4261, Jul. 2020.

- [28] A. Papazafeiropoulos, P. Kourtessis, M. Di Renzo, S. Chatzinotas, and J. M. Senior, "Performance analysis of cell-free massive MIMO systems: A stochastic geometry approach," *IEEE Trans. Veh. Technol.*, vol. 69, no. 4, pp. 3523–3537, Apr. 2020.
- [29] Y. Zhang, B. Di, H. Zhang, J. Lin, C. Xu, D. Zhang, Y. Li, and L. Song, "Beyond cell-free MIMO: Energy efficient reconfigurable intelligent surface aided cell-free MIMO communications," *IEEE Trans. Cogn. Commun. Netw.*, vol. 7, no. 2, pp. 412–426, Jun. 2021.
- [30] A. Lancho, G. Durisi, and L. Sanguinetti, "Cell-free massive MIMO with short packets," in *Proc. IEEE Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Sept. 2021, pp. 416–420.
- [31] A. A. Nasir, H. D. Tuan, H. Q. Ngo, T. Q. Duong, and H. V. Poor, "Cell-free massive MIMO in the short blocklength regime for URLLC," *IEEE Trans. Wireless Commun.*, vol. 20, no. 9, pp. 5861–5871, Sept. 2021.
- [32] Z. Chen and E. Björnson, "Channel hardening and favorable propagation in cell-free massive MIMO with stochastic geometry," *IEEE Trans. Commun.*, vol. 66, no. 11, pp. 5205–5219, Nov. 2018.
- [33] O. Mehanna, K. Huang, B. Gopalakrishnan, A. Konar, and N. D. Sidiropoulos, "Feasible point pursuit and successive approximation of non-convex QCQPs," *IEEE Signal Process. Lett.*, vol. 22, no. 7, pp. 804–808, 2014.
- [34] T. Van Chien, E. Björnson, and E. G. Larsson, "Joint pilot design and uplink power allocation in multi-cell massive MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 17, no. 3, pp. 2000–2015, Mar. 2018.
- [35] T. C. Mai, H. Q. Ngo, M. Egan, and T. Q. Duong, "Pilot power control for cell-free massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 67, no. 11, pp. 11 264–11 268, Nov. 2018.
- [36] J. Zhang, E. Björnson, M. Matthaiou, D. W. K. Ng, H. Yang, and D. J. Love, "Prospective multiple antenna technologies for beyond 5G," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1637–1660, Aug. 2020.
- [37] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, Jul. 1948.
- [38] H. Ren, C. Pan, K. Wang, Y. Deng, M. ElKashlan, and A. Nallanathan, "Achievable data rate for URLLC-enabled UAV systems with 3-D channel model," *IEEE Wireless Commun. Lett.*, vol. 8, no. 6, pp. 1587–1590, Dec. 2019.
- [39] J. Zhang, Y. Wei, E. Björnson, Y. Han, and S. Jin, "Performance analysis and power control of cell-free massive MIMO systems with hardware impairments," *IEEE Access*, vol. 6, pp. 55 302–55 314, 2018.
- [40] L. Luo, J. Zhang, S. Chen, X. Zhang, B. Ai, and D. W. K. Ng, "Downlink power control for cell-free massive MIMO with deep reinforcement learning," *IEEE Trans. Veh. Tech.*, Jun. 2022.
- [41] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optim. Eng.*, vol. 8, no. 1, pp. 67–127, May 2007.
- [42] C. Pan, H. Zhu, N. J. Gomes, and J. Wang, "Joint user selection and energy minimization for ultra-dense multi-channel C-RAN with incomplete CSI," *IEEE J. Sel. Areas Commun.*, vol. 35, no. 8, pp. 1809–1824, Aug. 2017.
- [43] J. Zhang, J. Zhang, D. W. K. Ng, S. Jin, and B. Ai, "Improving sum-rate of cell-free massive MIMO with expanded compute-and-forward," *IEEE Trans. Signal Process.*, vol. 70, pp. 202–215, 2021.
- [44] Z. Wang, J. Zhang, B. Ai, C. Yuen, and M. Debbah, "Uplink performance of cell-free massive MIMO with multi-antenna users over jointly-correlated Rayleigh fading channels," *IEEE Trans. Wireless Commun.*, vol. 21, no. 9, pp. 7391–7406, Sept. 2022.
- [45] H. Q. Ngo, A. Ashikhmin, H. Yang, E. G. Larsson, and T. L. Marzetta, "Cell-free massive MIMO versus small cells," *IEEE Trans. Wireless Commun.*, vol. 16, no. 3, pp. 1834–1850, Mar. 2017.
- [46] A. M. Tulino and S. Verdú, "Random matrix theory and wireless communications," *Found. Trends Commun. Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.