Joint Pilot and Payload Power Allocation for Massive-MIMO-enabled URLLC IIoT Networks

Hong Ren, Cunhua Pan, Yansha Deng, Maged Elkashlan, Arumugam
Nallanathan, *Fellow, IEEE*

Abstract

The Fourth Industrial Revolution (Industrial 4.0) is coming, and this revolution will fundamentally enhance the way the factories manufacture products. The conventional wired lines connecting central controller to robots or actuators will be replaced by wireless communication networks due to its low cost of maintenance and high deployment flexibility. However, some critical industrial applications require ultra-high reliability and low latency communication (URLLC). In this paper, we advocate the adoption of massive multiple-input multiple output (MIMO) to support the wireless transmission for industrial applications as it can provide deterministic communications similar as wired lines thanks to its channel hardening effects. To reduce the latency, the channel blocklength for packet transmission is finite, and suffers from transmission rate degradation and decoding error probability. Thus, conventional resource allocation for massive MIMO transmission based on Shannon capacity assuming the infinite channel blocklength is no longer optimal. We first derive the closed-form expression of lower bound (LB) of achievable uplink data rate for massive MIMO system with imperfect channel state information (CSI) for both maximum-ratio combining (MRC) and zero-forcing (ZF) receivers. Then, we propose novel low-complexity algorithms to solve the achievable data rate maximization problems by jointly optimizing the pilot and payload transmission power for both MRC and ZF. Simulation results confirm the rapid convergence speed and performance advantage over the existing benchmark algorithms.

Index Terms

URLLC, Industrial 4.0, Industrial Internet-of-Things (IIoT), Massive MIMO.

H. Ren, C. Pan, M. Elkashlan and A. Nallanathan are with School of Electronic Engineering and Computer Science, Queen Mary University of London, London, E1 4NS, U.K. (Email:h.ren, c.pan, a.nallanathan@qmul.ac.uk). Y. Deng is with the Department of Informatics, King's College London, London WC2R 2LS, U.K. (e-mail:yansha.deng@kcl.ac.uk).

I. INTRODUCTION

Industry 4.0 has been envisioned as the future paradigm for the next generation of industrial systems, which integrates advanced manufacturing functions with the industrial internet-of-things (IIoT) to create a more intelligent and automatic digital manufacturing system [1]. Traditionally, industrial control systems mainly rely on wired connections such as cables or optical fiber, since the current wireless networks cannot meet their stringent latency and reliability requirements. However, there are some drawbacks to deploying wired lines. First, significant cost will be incurred by the installation and maintenance. Second, wired lines are vulnerable to wear and tear in motion control applications, and suffer from aging. Finally, they cannot be deployed in some harsh environments, such as those with high temperatures and rotating part. Hence, to make Industry 4.0 a reality, it is imperative to design wireless networks tailored for industrial applications to replace the traditional wired lines. Typical industrial applications require deterministic communications with ultra reliability ($1 - 10^{-9}$) and low latency (1 ms), such as factory automation (FA), power system protection (PSP), and power electronics control (PEC) [2]. Significant research efforts have been devoted to the design of wireless communications in industrial applications. However, most of the existing works mainly focused on the adaption of the upper layers of conventional wireless networks to achieve deterministic communications, while keeping the physical layer untouched. Some related standards are WirelessHART, wireless interface for sensors and actuators (WISA), and Wireless Networks for Industrial Automation/Process Automation (WIA-PA) [3]. Although keeping the wireless standards of physical layers can allow faster design and better compatibility, it leads to a fundamental bottleneck for the system performance. None of the above standards can meet the stringent demand requested by the most critical FA, PSP and PEC applications. As a result, more efforts should be devoted to the design from the physical layer perspective of view.

From the physical layer perspective, the dominate feature of the industrial applications is that the packet transmission should be completed within short blocklength due to low latency requirement [4]. Hence, the transmission is not error-free with any finite/short blocklength channel codes. In this case, Shannon capacity formula is not applicable since it is based on the principles of the law of large numbers, and we need to design the resource allocation by considering the decoding error probability requirement. In [5], Peter *et al.* have derived the

approximation formula of the maximum achievable data rate with finite blocklength transmission, which characterises the complicated relationships among decoding error probability, channel blocklength, and signal-to-noise ratio (SNR). Unlike Shannon capacity formula, the data rate expression under short packet transmission is neither convex nor concave with respect to the SNR or blocklength [6]. As a result, the optimal resource allocation under this formula is difficult to obtain.

Recently, there are increasing research studying the transmission design based on the short packet transmission capacity formula [7]–[13]. In specific, the effective throughput maximization was studied in [7] for a two-device downlink non-orthogonal multiple access (NOMA) system. The overall error probability is minimized in [8] for a simultaneous wireless information and power transfer (SWIPT)-enabled decode-and-forward (DF) relaying network. The decoding error probability minimization was investigated in [9] for a unmanned aerial vehicle (UAV)-enabled DF relay system. We recently proposed a low-complexity power and blocklength optimization algorithm for both orthogonal multiple access (OMA) and NOMA in a two-hop relay system in [10]. However, all these contributions are limited to a simple scenario with two devices, where one device acts as a relay and the other as the destination node. In industrial applications, the central controller needs to support a large number of devices [3]. On one hand, Chen *et al.* in [11] investigated the effective capacity maximization problem for wireless-powered IoT network with multiple devices operating in a time division multiple access (TDMA) mode. However, the time budget is already tight, and the portion allocated to each device will be marginal. Hence, TDMA strategy is not suitable for the applications in industrial applications with extremely stringent latency target. On the other hand, the authors in [12] [13] studied the resource allocation for multiple devices operating under the orthogonal frequency division multiple access (OFDMA) mode. Unfortunately, this requires huge amount of system bandwidth, which is not feasible as some industrial applications operate over unlicensed spectrum [14].

By equipping a large number of antennas at the base station (BS), massive multiple-input multiple-output (MIMO) has been widely regarded as the key enabler for the creation of the fifth generation (5G) wireless networks [15]. By exploiting excessive number of spatial degrees of freedom, massive MIMO is capable of supporting multiple devices simultaneously without additional time or frequency resources. In addition, due to the channel hardening effect, massive MIMO is more immune to the fast fading and can provide deterministic communications required

by the industrial applications. Due to these attractive advantages, massive MIMO is ideal for supporting industrial applications with stringent quality of services (QoS) requirements. However, most of the existing literature adopted Shannon capacity as the performance metric to optimize the resource allocation [16]–[18], which implicitly assumes the infinite channel blocklength. Therefore, conventional resource allocation solution based on Shannon capacity is not optimal for industrial applications with short channel blocklength. To the best of our knowledge, we are the first to study the resource allocation for massive MIMO providing ultra-reliability and low-latency communications (URLLC) for any number of devices. Specifically, our contributions are summarized as follows:

1) We derive the closed-form lower bounds (LBs) on the achievable rates for a uplink massive MIMO system by considering the imperfect channel state information (CSI) with finite channel blocklength, and assuming both maximum-ratio combining (MRC) and zero-forcing (ZF) receivers. They can be regarded as the conventional Shannon capacity minus a penalty term due to short packet transmission. Simulation results confirm the tightness of the LBs. Given fixed delay budget, we formulate an optimization problem with the objective to maximize the weighted sum rate by jointly optimizing the pilot and payload transmission power subject to the decoding error probability, the minimum data rates, and the energy constraints for all URLLC devices.

2) For the case with MRC receiver, the formulated optimization problem is non-convex due to the complicated expression of data rate LBs, and it is difficult to find the globally optimal solution. To deal with this issue, we first approximate the penalty term in the data LB as a log-function, which can facilitate the transformation from the original optimization problem to a series of geometric programs (GPs). Each GP problem can be efficiently solved with polynomial time. Besides, we provide a novel method to check the feasibility of the original problem, and provide both complexity and convergence analysis.

3) For the case with ZF receiver, the data rate LB is more complicated and the algorithm proposed for MRC cannot be directly applied since the nominator of the signal-to-interference-plus-noise (SINR) is a posynomial function. To handle this issue, we approximate the polynomial functions with their best local monomial approximations and then transform the optimization into a series of GPs with low-complexity. Convergence analysis tailored

for the ZF receiver is further provided.

4) Simulation results show that our proposed algorithms converge rapidly, which verifies the low-complexity of the proposed algorithm. In addition, it is also shown that the proposed algorithm outperforms the benchmark schemes, especially the ones adopting Shannon capacity as the optimization performance metric, which emphasizes the importance of using short packet transmission theory.

The remainder of this paper is organized as follows. In Section II, system model and problem formulation are provided. In Section III, we provide a low-complexity algorithm for joint design of pilot and payload power allocation for the MRC case. The ZF case is studied in Section IV. Then, simulation results and analysis are presented in Section V. The final conclusion is drawn in Section VI.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Factory System Model

Consider a uplink multi-device Massive MIMO communication in one factory as shown in Fig. 1, where the central controller (CC) serves K devices, e.g., actuator, robot, etc. The devices need to send their emergency information of URLLC requirements such as measured data or their current operation states to the CC. Thus, the CC can process these data information immediately and provide prompt response/feedback. For simplicity, we focus on the uplink transmission and the solutions for downlink can be similarly derived. The set of the devices is denoted as $\mathcal{K} = \{1, 2, \cdots, K\}$. The CC is equipped with M antennas and each device is equipped with single antenna due to their low signal processing capability, where $M \gg K$. Let us denote $\mathbf{h}_k \in \mathbb{C}^{M \times 1}$ as the channel vector from the CC to the k-th device and can be decomposed as $\mathbf{h}_k = \sqrt{\alpha_k} \bar{\mathbf{h}}_k$, where α_k denotes the large-scale channel gain that includes the pathloss and shadowing, and $\bar{\mathbf{h}}_k$ denotes the small-scale fading following the distribution of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. Let us denote $\mathbf{H} \in \mathbb{C}^{M \times K}$ as the channel matrix from the K devices to the CC, with $\mathbf{H} = [\mathbf{h}_1, \mathbf{h}_2, \cdots, \mathbf{h}_K]$.

The K devices need to transmit K packets to the controller. Then, the $M \times 1$ received signal vector at the CC is given by

$$\mathbf{y} = \sum\nolimits_{k \in \mathcal{K}} \mathbf{h}_k \sqrt{p_k^d} s_k + \mathbf{n}, \tag{1}$$

where p_k^d is the data transmission power of the kth device, s_k is the zero mean and unit variance Gaussian information message from the kth device, and $\mathbf{n} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_M)$ is the additive noise

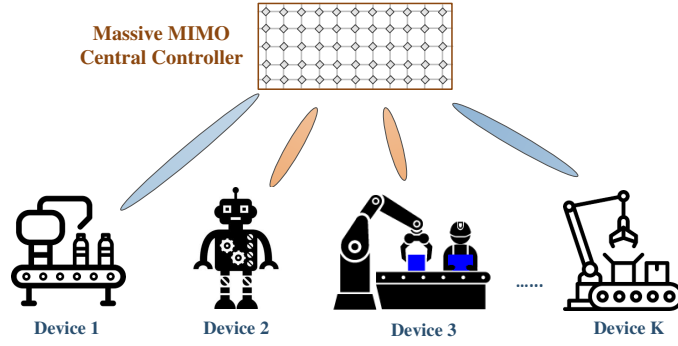during the data transmission, where the variance of each element is normalized to unit.

Fig. 1: Factory scenario where a massive MIMO central controller serves multiple devices

B. Channel Estimation in Massive MIMO URLLC

Due to the channel hardening [19] brought by massive MIMO, the system is more immune to the fast fading, which can provide high reliable services for the devices. However, to reap the benefits brought by massive MIMO, the CSI should be available at the CC. Furthermore, TDD mode is always taken as an enabler for massive MIMO systems since downlink instantaneous CSI is obtained by estimating uplink CSI based on channel reciprocity [20]. In addition, for machine type communications, the devices may not be able to perform complicated signal processing tasks required in frequency division duplexing (FDD) systems, such as channel estimation calculation, quantization, etc. More time slots are needed for CSI feedback. Hence, the TDD protocol is adopted in this paper. All the devices should be allocated with orthogonal pilot resources so that the CC is able to distinguish the channels from different devices, thus, the number of symbols for channel estimation should be no smaller than the number of devices [19].

In a massive MIMO URLLC scenario, each block mainly consists of two parts: 1) l_p symbols for channel estimation (the pilot sequence is of length l_p); 2) l_d symbols for the K devices' data transmission, thus the total number of symbols of the frame is denoted as $L = l_p + l_d$. As the symbols used for data transmission is already limited, we assume that K devices sharing the same symbol duration and the frame structure is illustrated in Fig. 2. Accordingly, the time durations for channel estimation and data transmission in one frame are given by $t_p = l_p/B$ and $t_d = l_d/B$, respectively, where B is the bandwidth of the system.

In the training phase, all devices simultaneously and synchronously transmit orthogonal pilot sequences $\mathbf{q}_1, \cdots, \mathbf{q}_{l_p} \in \mathbb{C}^{l_p \times 1}$ to the CC, with $\mathbf{q}_k^H \mathbf{q}_k = 1$ and $\mathbf{q}_i^H \mathbf{q}_j = 0$, $i \neq j$. Hence, the
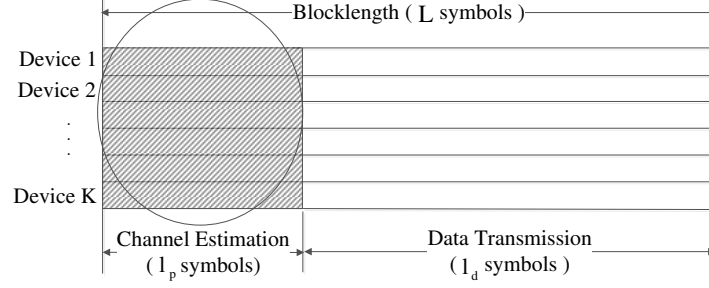
Fig. 2: Block diagram scheme in TDD URLLC Massive MIMO scheme

minimum length of the pilot sequences to guarantee the orthogonality is equal to $l_p = K$. Based on the received signal, the CC estimates the channel conditions of all devices, and the received pilot signal at the CC is

$$\mathbf{Y}^p = \sum_{k \in \mathcal{K}} \sqrt{K p_k^p} \mathbf{h}_k \mathbf{q}_k^H + \mathbf{N}, \qquad (2)$$

where p_k^p is the pilot transmit power at the kth device, and $\mathbf{N} \in \mathbb{C}^{M \times K}$ is the additive Gaussian noise matrix received during the training phase, whose elements are independently generated and follow the distribution of $\mathcal{CN}(0, 1)$. To obtain channel $\mathbf{h}_k$, the CC first multiplies $\mathbf{Y}^p$ by $\frac{1}{\sqrt{K p_k^p}} \mathbf{q}_k$, which yields

$$\mathbf{y}_k^p = \frac{1}{\sqrt{K p_k^p}} \mathbf{Y}^p \mathbf{q}_k = \mathbf{h}_k + \mathbf{n}_k^p, \qquad (3)$$

where $\mathbf{n}_k^p = \frac{1}{\sqrt{K p_k^p}} \mathbf{N} \mathbf{q}_k$. Since $\mathbf{q}_k$ is a unit-norm vector, it is easy to show that $\mathbf{n}_k^p$ is still Gaussian distribution whose elements are independently and identically distributed as $\mathcal{CN}(0, \frac{1}{K p_k^p} \mathbf{I}_M)$. The MMSE estimate of channel $\mathbf{h}_k$ is given by

$$\hat{\mathbf{h}}_k = \frac{\alpha_k K p_k^p}{\alpha_k K p_k^p + 1} \mathbf{y}_k^p, \qquad (4)$$

which follows the distribution of $\mathcal{CN}(\mathbf{0}, \sigma_k \mathbf{I})$ with σ_k given by

$$\sigma_k = \frac{\alpha_k^2 K p_k^p}{\alpha_k K p_k^p + 1}. \qquad (5)$$

According to the property of MMSE estimation, channel estimation error $\tilde{\mathbf{h}}_k = \mathbf{h}_k - \hat{\mathbf{h}}_k$ is independent of $\hat{\mathbf{h}}_k$, and follows the distribution of $\mathcal{CN}(\mathbf{0}, \delta_k \mathbf{I}_M)$, where δ_k is given by

$$\delta_k = \frac{\alpha_k}{\alpha_k K p_k^p + 1}. \qquad (6)$$

C. Achievable Data Rate for Massive MIMO URRLC

By taking into account the number of symbols for pilot transmission, to achieve the decoding error probability of ε_k for the kth device, the instantaneous achievable data rate R_k can be

accurately approximated by [21]

$$R_k \approx (1 - \beta) \log_2(1 + \gamma_k) - \sqrt{\frac{(1 - \beta) V_k}{L}} \frac{Q^{-1}(\varepsilon_k)}{\ln 2}, \tag{7}$$

where β is equal to $\beta = K/L$, γ_k is the signal to interference plus noise ratio (SINR) of the kth device, Q^{-1} is the inverse function $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-\frac{t^2}{2}} \mathrm{d}t$, and V_k is the channel dispersion given by $V_k = 1 - (1 + \gamma_k)^{-2}$. As seen from (7), when the blocklength L approaches infinity, the data rate R_k will approach $(1 - \beta) \log_2(1 + \gamma_k)$, which is the classic Shannon capacity. The second term in (7) can be interpreted as a penalty on the rate in order to guarantee the decoding error probability ε_k.

In the following, we derive the expression of γ_k for two different low-complexity detection schemes: 1) maximum-ratio combining (MRC); 2) zero-forcing (ZF).

Define the estimated channels as $\widehat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \cdots, \hat{\mathbf{h}}_K]$ and channel estimation errors as $\widetilde{\mathbf{H}} = [\tilde{\mathbf{h}}_1, \tilde{\mathbf{h}}_2, \cdots, \tilde{\mathbf{h}}_K]$. Let $\mathbf{A}$ be an $M \times K$ linear detection matrix that is based on the estimated channel $\widehat{\mathbf{H}}$. By using the linear detection $\mathbf{A}$, the received signal can be processed as

$$\mathbf{y}^D = \mathbf{A}^H \mathbf{y}. \tag{8}$$

Two conventional low-complexity linear detectors are considered:

$$\mathbf{A} = \begin{cases} \widehat{\mathbf{H}}, & \text{for} \quad \text{MRC} \\ \widehat{\mathbf{H}} \left(\widehat{\mathbf{H}}^H \widehat{\mathbf{H}} \right)^{-1}, & \text{for} \quad \text{ZF}. \end{cases} \tag{9}$$

Then, the processed signal after using the detector is given by

$$\begin{aligned} \mathbf{y}^D &= \mathbf{A}^H \sum_{i \in \mathcal{K}} \mathbf{h}_i \sqrt{p_i^d} s_i + \mathbf{A}^H \mathbf{n}. \\ &= \mathbf{A}^H \sum_{i \in \mathcal{K}} \hat{\mathbf{h}}_i \sqrt{p_i^d} s_i + \mathbf{A}^H \sum_{i \in \mathcal{K}} \tilde{\mathbf{h}}_i \sqrt{p_i^d} s_i + \mathbf{A}^H \mathbf{n}, \end{aligned} \tag{10}$$

where the last equality is obtained by using $\mathbf{h}_k = \hat{\mathbf{h}}_k + \tilde{\mathbf{h}}_k$. The detection signal for the kth device is given by

$$\mathbf{y}_k^D = \mathbf{a}_k^H \hat{\mathbf{h}}_k \sqrt{p_k^d} s_k + \mathbf{a}_k^H \sum_{i \in \mathcal{K} \backslash k} \hat{\mathbf{h}}_i \sqrt{p_i^d} s_i + \mathbf{a}_k^H \sum_{i \in \mathcal{K}} \tilde{\mathbf{h}}_i \sqrt{p_i^d} s_i + \mathbf{a}_k^H \mathbf{n}. \tag{11}$$

where $\mathbf{a}_k$ is the kth column of matrix $\mathbf{A}$. Since $\widehat{\mathbf{H}}$ and $\widetilde{\mathbf{H}}$ are independent, $\mathbf{a}_k$ is also independent of $\widetilde{\mathbf{H}}$. The CC will treat the estimated channel as the true channel, and the last three terms of
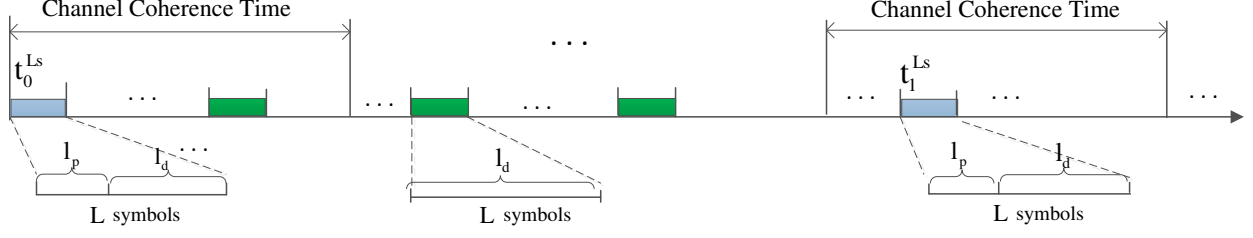
Fig. 3: Block diagram for transmission in TDD URLLC Massive MIMO scheme.

(11) are regarded as interference and noise. Then, the SINR for the kth device γ_k is given by

$$
\gamma_k = \frac{p_k^d \left| \mathbf{a}_k^H \hat{\mathbf{h}}_k \right|^2}{\sum_{i \in \mathcal{K} \backslash k} p_i^d \left| \mathbf{a}_k^H \hat{\mathbf{h}}_i \right|^2 + \sum_{i \in \mathcal{K}} p_i^d \left| \mathbf{a}_k^H \tilde{\mathbf{h}}_i \right|^2 + \|\mathbf{a}_k\|^2}.
\tag{12}
$$

Remark: Massive MIMO can offer the channel hardening effect, where the channel variations between different channel fading blocks can be averaged out and the achievable data rate for these fading blocks mainly depend on the large-scale fading, which changes very slowly. As a result, we first derive the lower bound (LB) for the achievable data rate as a function of large-scale channel fading parameters, and then optimize the resource allocation based on the large-scale fading information rather than the small-scale fading, which can significantly *reduce the computational delay* that is beneficial for URLLC applications. In other words, when the large-scale fading parameters of all devices are given, we can use our developed algorithm to find the optimal power allocation, which can be used for consecutive channel fading blocks. The algorithms are needed to be rerun only when the large-scale channel fading parameters have changed, which vary much slowly compared with the small-scale channel fading. To make it more clear, a block diagram scheme is given in Fig. 3, where the large-scale channel gains of any devices vary at $t = t_0^{Ls}, t_1^{Ls}$. In general, the channel coherence time is much longer than the packet transmission time as we consider the URLLC services, and the time difference for large-scale fading ($t_1^{Ls} - t_0^{Ls}$) is much larger than the channel coherence time. The power allocation obtained at time $t = t_0^{Ls}$ can be employed for the subsequent transmissions until $t = t_1^{Ls}$, which significantly reduce the computational time. Please note in our scheme, the channel estimation should be performed at the beginning of each channel coherence since the decoding requires the channel state information as shown in (9), while the power allocation needs to be updated once the large-scale fading gains change.

Due to the channel hardening effect, in this paper we focus on the ergodic achievable data

rate that is defined as $\bar{R}_k = \mathbb{E}\{R_k\}$, where the expectation is taken over the randomness of $\left\{\hat{\mathbf{h}}_k, \tilde{\mathbf{h}}_k, \forall k\right\}$. Unfortunately, the exact average achievable data rate $\bar{R}_k$ with channel uncertainty is not available. In the following, we aim to derive the closed-form expression of the LB of the rate expression, which is more tractable to analyse and optimize. To this end, we first define function $f(x)$ as

$$f(x) = \ln\left(1 + \frac{1}{x}\right) - a\sqrt{\frac{2x+1}{(x+1)^2}} \geq 0, x > 0 \tag{13}$$

where a is a fixed positive value. In the following, we derive the feasible region of function $f(x)$. Since $f(x) \geq 0$, from (13) we have

$$a \leq \frac{(x+1)\ln\left(1 + \frac{1}{x}\right)}{\sqrt{2x+1}} \triangleq g(x). \tag{14}$$

The first-order derivative of $g(x)$ with respect to x is given by

$$g'(x) = \frac{-2 - \frac{1}{x} + x\ln\left(1 + \frac{1}{x}\right)}{(2x+1)^{\frac{3}{2}}} \leq \frac{-1 - \frac{1}{x}}{(2x+1)^{\frac{3}{2}}} < 0 \tag{15}$$

where the first inequality follows by using the relation $\ln\left(1 + \frac{1}{x}\right) < \frac{1}{x}$. Hence, $g(x)$ is a monotonically decreasing function of x. In addition, $\lim_{x\to 0} g(x) = \infty$ and $\lim_{x\to\infty} g(x) = 0$, where the latter equation is obtained by using the L'Hospital's rule. Hence, from (14), we know that the feasible region of $f(x)$ is given by

$$\Omega = \left\{x \mid 0 < x \leq g^{-1}(a)\right\}. \tag{16}$$

Then, we have the following lemma.

Lemma 1: Function $f(x)$ defined in (13) is decreasing and convex for $x \in \Omega$.

Proof: Please refer to Appendix B in our recent work [22]. ∎

Based on Lemma 1, we are able to derive the LB of $\bar{R}_k$ in the following. The instantaneous data rate R_k in (7) can be written as follows:

$$R_k = \frac{1 - \beta}{\ln 2} f_k\left(\frac{1}{\gamma_k}\right), \tag{17}$$

where function $f_k(\cdot)$ is in the same format as $f(\cdot)$ in (13), where the parameter a is

$$a_k = Q^{-1}(\varepsilon_k)/\sqrt{L(1-\beta)}. \tag{18}$$

In addition, we need to guarantee that $R_k \geq 0$, and thus $\gamma_k \geq 1/g^{-1}(a_k)$. By using lemma 1 and Jensen's inequality, we obtain the following LB on the ergodic data rate:

$$\bar{R}_k \geq \hat{R}_k \triangleq \frac{1-\beta}{\ln 2} f_k \left(\mathbb{E} \left\{ \frac{1}{\gamma_k} \right\} \right). \tag{19}$$

In the following theorem, we derive the expression of $\hat{R}_k$ for each beamforming solution.

Theorem 1: The ergodic achievable rate for the kth device for MRC in finite blocklenghth regime can be lower bounded by:

$$\hat{R}_k \triangleq \frac{1-\beta}{\ln 2} f_k \left(\frac{1}{\hat{\gamma}_k} \right) \tag{20}$$

where $\hat{\gamma}_k$ is given by

$$\hat{\gamma}_k = \frac{p_k^d (M-1) \sigma_k}{\sum_{i \in \mathcal{K} \backslash k} p_i^d \sigma_i + \sum_{i \in \mathcal{K}} p_i^d \delta_i + 1}. \tag{21}$$

Proof: Please refer to Appendix A. ∎

For the ZF detection, the LB of the ergodic data rate is given by the following theorem.

Theorem 2: The ergodic achievable rate for the kth device for ZF in finite blocklength regime is lower bounded by:

$$\hat{R}_k \triangleq \frac{1-\beta}{\ln 2} f_k \left(\frac{1}{\hat{\gamma}_k} \right) \tag{22}$$

where $\hat{\gamma}_k$ is given by

$$\hat{\gamma}_k = \frac{(M-K) \sigma_k p_k^d}{\sum_{i \in \mathcal{K}} p_i^d \delta_i + 1}. \tag{23}$$

Proof: Please refer to Appendix B. ∎

In the proof of Theorem 1 and Theorem 2, we notice that the above derived LB of the ergodic data rates are valid for any number of antennas ($M > K$ for the ZF). However, the gap between the LB and the actual ergodic data rate is reduced when the number of antennas is large, which is the case for massive MIMO system. These LBs are commonly used in the literature concerning massive MIMO systems. Hence, we use these lower bounds throughout this paper. In the following, we aim to optimize the pilot power p_k^p and data transmission power p_k^d to maximize the weighted sum rate. The optimization is performed only when any large-scale fading parameter changes. The simulation results in Section V also verify the tightness of the derived LBs. Hence, the optimization solutions are applicable for *a long time scale*, which is appealing for URLLC applications.

D. Problem Formulation

In this paper, we jointly optimize the power allocation for pilot and data transmission of each device for maximizing the weighted sum rate of all devices. By using Theorem 1, the weighted

sum rate maximization problem can be formulated as

$$\max_{\{p_k^p\},\{p_k^d\}} \quad \sum_{k\in\mathcal{K}} w_k \hat{R}_k \tag{24a}$$

$$\text{s.t.} \quad \hat{R}_k \geq R_k^{\text{req}}, \forall k, \tag{24b}$$

$$K p_k^p + (L-K) p_k^d \leq E_k, \forall k, \tag{24c}$$

where $\hat{R}_k$ and $\hat{\gamma}_k$ are given in (20) and (21) for MRC and (22) and (23) for ZF, respectively, w_k is the weight of device k used to guarantee the fairness among the devices, constraint (24b) denotes the minimum data rate requirement for the k-th device, constraint (24c) means the energy constraint for each device.

Power control for weighted sum rate problem with interference is well known to be NP-hard problem even under perfect CSI [23]. It becomes more complicated for the more general case with imperfect CSI and finite blocklength. In this paper, we aim for designing efficient algorithms with polynomial-time complexity to solve the weighted sum rate problem with imperfect CSI.

To this end, we first simplify the problem formulation in (24). The first-order derivative of $\hat{R}_k$ w.r.t. $\hat{\gamma}_k$ is given by $\hat{R}_k' = -\frac{(1-\beta)}{\hat{\gamma}_k^2 \ln 2} f_k'\left(\frac{1}{\hat{\gamma}_k}\right) \geq 0$ [1]. Hence, constraint (24b) can be transformed as

$$\hat{\gamma}_k \geq 1 \Big/ f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1-\beta}\right), \forall k. \tag{25}$$

To additionally simplify the problem formulation in (24), we introduce auxiliary variables $\chi_k, \forall k$, and then Problem (24) can be equivalently transformed as follows

$$\max_{\{\chi_k\},\{p_k^p\},\{p_k^d\}} \quad \sum_{k\in\mathcal{K}} \tilde{w}_k \left[\ln(1+\chi_k) - a_k G(\chi_k)\right] \tag{26a}$$

$$\text{s.t.} \quad \hat{\gamma}_k \geq \chi_k, \forall k, \tag{26b}$$

$$(25), (24c), \tag{26c}$$

where $\tilde{w}_k = \frac{(1-\beta)w_k}{\ln 2}$ and $G(\chi_k) = \sqrt{1 - \frac{1}{(1+\chi_k)^2}}$. The equivalence between Problem (26) and Problem (24) can be readily proved by using the contradiction method. Problem (26) and Problem (24) are equivalent in the sense that they have the same power allocation solutions and the same objective function (OF) value. Hence, in the following, we focus on solving Problem (26).

Due to different SINR expressions, in the following two sections, we optimize the power allocation for MRC and ZF, respectively.

---

[1] Since $\hat{R}_k \geq R_k^{\text{req}} > 0$, the feasible $\hat{\gamma}_k$ must lie in the range of $\mathbf{\Omega} = \left\{\hat{\gamma}_k | 0 < 1/\hat{\gamma}_k \leq g^{-1}(a_k)\right\}$. Hence, Lemma 1 holds and we have $f_k'\left(\frac{1}{\hat{\gamma}_k}\right) \geq 0$.

III. Weighted Sum Data Rate for MRC

In this section, we aim to deal with weighted sum rate maximization problem for MRC.

A. Algorithm Design

The complicated function $G(\chi_k)$ in (26a) makes the optimization problem difficult to solve. To resolve this issue, we first study several properties of this function.

Lemma 2: $G(x)$ is a concave function of x.

Proof: Please refer to Appendix C. ■

For the URLLC applications, the decoding error probability ε_k for each device is much smaller than 0.5 so that a_k is a strictly positive value, and the OF of Problem (26) is to maximize the difference of two concave functions. However, due to the non-convex constraints in Problem (26). This problem does not belong to the class of difference of convex (DC) problem. In addition, due to the additional term of $a_k G(\chi_k)$ in the OF of Problem (26), this problem cannot be solved by dealing with a sequence of geometric programs (GPs) as in [24], which considered Shannon capacity formula under the infinite blocklength regime. The intuitive method to solve Problem (26) is to approximate function $G(\chi_k)$ as its first-order Taylor series expansion and solve the approximated problem until convergence. However, the approximation is in linear form while the first term in the OF is in log-function form. The successive GP method in [24] is still not applicable. To deal with this issue, we approximate function $G(x)$ as a log-function as shown in the following lemma.

Lemma 3: For any given $\tilde{x} \geq \frac{\sqrt{17}-3}{4}$, the following inequality holds:

$$G(x) = \sqrt{1 - \frac{1}{(1+x)^2}} \leq \rho\ln(x) + \eta \triangleq F(x), \forall x \geq \frac{\sqrt{17}-3}{4}, \tag{27}$$

where ρ and η are given by

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

and

$$\eta = \sqrt{1 - \frac{1}{(1+\tilde{x})^2}} - \rho\ln(\tilde{x}). \tag{29}$$

In addition, we have:

$$G(\tilde{x}) = F(\tilde{x}), G'(\tilde{x}) = F'(\tilde{x}), \tag{30}$$

which means that the approximation $F(x)$ is tight at $x = \tilde{x}$.

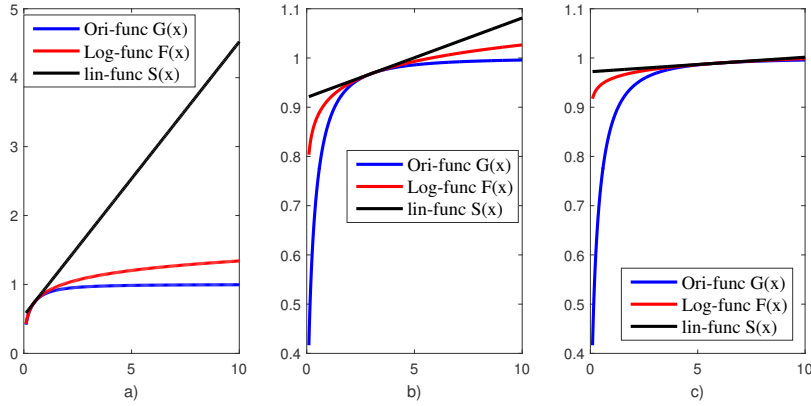***Proof***: Please refer to Appendix D. ■

Fig. 4: Approximation error for different approximation functions at three different points: a) $\tilde{x} = 0.5$; b) $\tilde{x} = 3$; c) $\tilde{x} = 6$.

According to (16), the optimization variable $\hat{\gamma}_k$ should be no smaller than $1/g^{-1}(a_k)$. Note that $1/g^{-1}(a_k)$ is a decreasing function of the decoding error probability and blocklength, while an increasing function of the number of device. Hence, for typical FA cases [25], where the typical required error probability is lower than 10^{-8}, the available channel blocklength is smaller than 200, and the number of devices that should be supported is larger than 5, $1/g^{-1}(a_k)$ is larger than $\frac{\sqrt{17}-3}{4}$, which implies that Lemma 3 is applicable for our considered optimization problem in (26).

In Fig. 2, we compare the approximation accuracy of function $F(x)$ and the linear approximation function $S(x)$ defined as follows:

$$G(x) \le G(\tilde{x}) + G'(\tilde{x})\left(x - \tilde{x}\right) \triangleq S(x), \tag{31}$$

where the inequality holds since $G(x)$ is a concave function proved in Lemma 2. It can be observed from Fig. 4 that the approximation function $F(x)$ is more accurate than the linear function $S(x)$ over the whole region of x at different points of $\tilde{x}$. We can also find that the curve of the log-function $F(x)$ is always above the curve of the linear function $S(x)$, which verifies the correctness of the theoretical conclusion in Lemma 3. The most important advantage of approximating $G(x)$ as the log-function $F(x)$ is that we can transform the original problem into a GP problems that enables us to find the optimal solution.

In the following lemma, we also provide the LB of $\ln(1 + \chi_k)$, which enables us to develop low-complexity algorithms.

Lemma 4: For any given $\tilde{x} \ge 0$, function $\ln(1 + x)$ is lower bounded by

$$\ln(1 + x) \geq \hat{\rho} \ln x + \hat{\eta}, \forall x \geq 0, \tag{32}$$

where $\hat{\rho}$ is given by

$$\hat{\rho} = \frac{\tilde{x}}{1 + \tilde{x}}, \hat{\eta} = \ln(1 + \tilde{x}) - \frac{\tilde{x}}{1 + \tilde{x}} \ln \tilde{x}. \tag{33}$$

In addition, the bound is tight at $x = \tilde{x}$.

Proof: The proof is similar to those in Lemma 3, and thus omitted for simplicity. ∎

Based on Lemma 3 and Lemma 4, we are now ready to solve Problem (26). The main idea is to approximate the OF of Problem (26) as the approximated functions provided in Lemma 3 and Lemma 4, and then solve the approximate problem in an iterative manner. In the following, we provide the details of the iterative algorithm.

First, we denote $\mathbf{P}^{(i)} = \{p_k^{p(i)}, p_k^{d(i)}, \forall k\}$ as the power allocation in the i-th iteration, and the corresponding χ_k is given by $\chi_k^{(i)}$. Then, in the $i + 1$-th iteration, we can approximate $G(\chi_k)$ around $\chi_k^{(i)}$ as function $F(\chi_k) = \rho_k^{(i)} \ln \chi_k + \eta_k^{(i)}$, where $\rho_k^{(i)}$ and $\eta_k^{(i)}$ are obtained from (28) and (29) respectively with $\tilde{x} = \chi_k^{(i)}$. In addition, we can approximate $\ln(1 + \chi_k)$ around $\chi_k^{(i)}$ as $\hat{\rho}_k^{(i)} \ln \chi_k + \hat{\eta}_k^{(i)}$, where $\hat{\rho}_k^{(i)}$ and $\hat{\eta}_k^{(i)}$ can be obtained from (33) with $\tilde{x} = \chi_k^{(i)}$. By substituting these approximations into (26a) and recalling a_k is a positive value, we can obtain the LB of the OF by using Lemma 3 and Lemma 4 as follows

$$\sum_{k \in \mathcal{K}} \tilde{w}_k \left[\ln(1 + \chi_k) - a_k G(\chi_k) \right] \geq \sum_{\mathcal{K}} \tilde{w}_k \left[\hat{\rho}_k^{(i)} \ln \chi_k + \hat{\eta}_k^{(i)} - a_k \rho_k^{(i)} \ln \chi_k - a_k \eta_k^{(i)} \right]. \tag{34}$$

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

Next, we optimize the power allocation to maximize the LB of the OF instead of maximizing (26a) directly. In specific, the LB maximization problem to be solved in the $i + 1$-th iteration is formulated as

$$\max_{\{\chi_k\}, \{p_k^p\}, \{p_k^d\}} \sum_{k \in \mathcal{K}} \hat{w}_k^{(i)} \ln \chi_k \tag{35a}$$

$$\text{s.t.} \quad (26b), (25), (24c). \tag{35b}$$

where $\hat{w}_k^{(i)} = \tilde{w}_k \hat{\rho}_k^{(i)} - a_k \tilde{w}_k \rho_k^{(i)}$ and the constant term in the OF is omitted.

Then, we can transform the above optimization into a GP problem as follows:

$$\max_{\{\chi_k\},\{p_k^p\},\{p_k^d\}} \quad \prod_{k\in\mathcal{K}} \chi_k^{\hat{w}_k^{(i)}} \tag{36a}$$

$$\text{s.t.} \quad \sum_{i\in\mathcal{K}\setminus k}\alpha_i\alpha_k K\chi_k p_k^p p_i^d + \sum_{i\in\mathcal{K}}\alpha_i\chi_k p_i^d + \chi_k\alpha_k Kp_k^p + \chi_k \leq (M-1)K\alpha_k^2 p_k^p p_k^d, \forall k, \tag{36b}$$

$$\chi_k \geq 1 \Big/ f_k^{-1}\left(\frac{R_k^{\text{req}}\ln 2}{1-\beta}\right), \forall k, (24c). \tag{36c}$$

Although GP is not a convex optimization problem, it can be equivalently transformed into a convex optimization problem by applying a logarithmic change of variables. Therefore, the globally optimal solution of Problem (36) can be obtained by using the interior-point method [26]. Some software packages that can solve GP problem are MOSEK package and CVX [27].

Based on the above discussion, the iterative algorithm to solve Problem (26) is given in Algorithm 1.

Algorithm 1 Iterative algorithm for solving Problem (26) for MRC

1: Initialize iteration number $i = 1$, error tolerance ξ. Initialize a feasible power allocation $\{p_k^{p(0)}, p_k^{d(0)}, \forall k\}$, calculate $\{\chi_k^{(0)}, \rho_1^{(0)}, \rho_2^{(0)}, \hat{w}_k^{(0)}, \forall k\}$, and calculate the OF of Problem (26), denoted as $\text{Obj}^{(0)}$.

2: With given $\{\chi_k^{(i-1)}, \rho_1^{(i-1)}, \rho_2^{(i-1)}, \hat{w}_k^{(i-1)}, \forall k\}$, solve Problem (36) by using the CVX package to obtain $\{p_k^{p(i)}, p_k^{d(i)}, \chi_k^{(i)}, \forall k\}$.

3: Update $\{\rho_k^{(i)}, \rho_k^{(i)}, \hat{w}_k^{(i)}, \forall k\}$;

4: Calculate the new OF $\text{Obj}^{(i)}$. If $\left|\text{Obj}^{(i)} - \text{Obj}^{(i-1)}\right|/\text{Obj}^{(i)} < \xi$, terminate. Otherwise, set $i \leftarrow i + 1$, go to step 2.

B. Algorithm Analysis

1) Initialization of Algorithm 1: As shown in Step 1 of Algorithm 1, one has to find a feasible initial power allocation in order to make the algorithm work. Note that randomly selecting a set of power allocation solutions that satisfy the per-device energy constraints may not satisfy their minimum SINR requirements. Hence, one has to carefully choose the initial power allocation. In the following, we provide one alternative method to find the initial power allocation solutions.

Inspired by the user selection problem formulation in [28], [29], we construct the following alternative optimization problem by introducing an auxiliary variable φ:

$$\max_{\varphi,\{p_k^p\},\{p_k^d\}} \quad \varphi \tag{37a}$$

$$\text{s.t.} \quad \hat{\gamma}_k \geq \varphi \Big/ f_k^{-1}\left(\frac{R_k^{\text{req}}\ln 2}{1-\beta}\right), \forall k, (24c). \tag{37b}$$

Obviously, Problem (37) is always feasible since at least $\{\varphi = 0, p_k^p = 0, p_k^d = 0, \forall k\}$ is a feasible solution. It can be readily verified that the original Problem (24) is feasible if the optimal $\varphi \geq 1$, and the output power allocation can be adopted as the initial input for Algorithm 1. In this paper, we assume that Problem (24) is always feasible and the optimal φ in Problem (37) is always no smaller than one. Problem (37) can also be transformed into a GP problem, where the globally optimal solution can be obtained. The details are omitted here due to the limited space.

2) Convergence Analysis: In this part, we analyze the convergence of Algorithm 1. In the following, we show that $\mathrm{Obj}^{(i)} \leq \mathrm{Obj}^{(i+1)}$.

Since $\{\chi_k^{(i+1)}, \forall k\}$ is the optimal solution of Problem (35) in the $i + 1$-th iteration, we have

$$
\begin{aligned}
&\sum_{k \in \mathcal{K}} \tilde{w}_k \left[\hat{\rho}_k^{(i)} \ln\left(\chi_k^{(i+1)}\right) + \hat{\eta}_k^{(i)} - a_k \rho_k^{(i)} \ln(\chi_k^{(i+1)}) - a_k \eta_k^{(i)} \right] \\
&\geq \sum_{k \in \mathcal{K}} \tilde{w}_k \left[\hat{\rho}_k^{(i)} \ln\left(\chi_k^{(i)}\right) + \hat{\eta}_k^{(i)} - a_k \rho_k^{(i)} \ln(\chi_k^{(i)}) - a_k \eta_k^{(i)} \right] = \mathrm{Obj}^{(i)}.
\end{aligned}
\tag{38}
$$

By using inequality (34) with $\chi_k = \chi_k^{(i+1)}$, we have

$$
\begin{aligned}
&\sum_{k \in \mathcal{K}} \tilde{w}_k \left[\ln(1 + \chi_k^{(i+1)}) - a_k G(\chi_k^{(i+1)}) \right] \\
&\geq \sum_{k \in \mathcal{K}} \tilde{w}_k \left[\hat{\rho}_k^{(i)} \ln(\chi_k^{(i+1)}) + \hat{\eta}_k^{(i)} - a_k \rho_k^{(i)} \ln(\chi_k^{(i+1)}) - a_k \eta_k^{(i)} \right].
\end{aligned}
\tag{39}
$$

By combining (38) and (39), we have

$$
\mathrm{Obj}^{(i+1)} \geq \sum_{k \in \mathcal{K}} \tilde{w}_k \left[\hat{\rho}_k^{(i)} \ln(\chi_k^{(i+1)}) + \hat{\eta}_k^{(i)} - a_k \rho_k^{(i)} \ln(\chi_k^{(i+1)}) - a_k \eta_k^{(i)} \right] \geq \mathrm{Obj}^{(i)}.
\tag{40}
$$

In addition, since each device has its own energy constraint, the OF value of Problem (24) has upper bound. As a result, Algorithm 1 is guaranteed to converge.

3) Solution Analysis: Since the original problem (24) is non-convex, the globally optimal solution cannot be obtained in general. However, by using the similar proof as in Appendix B in [29], we can prove that Algorithm 1 can converge to the Karush-Kuhn-Tucker (KKT) point of Problem (24). The converged solution only depends on the initial input of Algorithm 1. Simulation results show that the algorithm almost converges to the same solution with different initial input solutions.

4) Complexity Analysis: The main complexity mainly lies in solving a GP problem in each iteration. In [30], the authors claimed that the GP problem can be efficiently solved by using the standard interior point methods with a worst-case polynomial-time complexity. The upper bound of the total number of Newton steps in the interior point method does not depend on

the number of variables, or the number of constraints. The derived upper bound shows that the barrier method converges linearly. By carefully choosing the parameters, the bound on the number of Newton steps can grow as $\sqrt{m}$ instead of m, where m is the number of constraints [31]. In addition, simulation results show that Algorithm 1 converges rapidly, which means that Algorithm 1 can converge to a local optimal solution with a polynomial time complexity.

IV. WEIGHTED SUM DATA RATE FOR ZF

In this section, we aim to deal with weighted sum rate maximization problem for ZF. Due to different expressions of the SINR of MRC and ZF, some derivations in MRC cannot be directly applied in the ZF case. In the following, we develop an efficient algorithm to solve Problem (26) for the ZF case.

A. Algorithm Design

By substituting (5) and (6) into (23), the SINR in the ZF case can be reformulated as

$$\hat{\gamma}_k = \frac{(M-K)\alpha_k^2 K p_k^p p_k^d \prod_{i\in\mathcal{K}} \left(1 + \alpha_i K p_i^p\right)}{\left(1 + \alpha_k K p_k^p\right)\left(\sum_{i\in\mathcal{K}} p_i^d \alpha_i \prod_{j\neq i} \left(1 + \alpha_j K p_j^p\right) + \prod_{i\in\mathcal{K}}\left(1 + \alpha_i K p_i^p\right)\right)}, \forall k. \qquad (41)$$

Unfortunately, since the nominator of $\hat{\gamma}_k$ in (41) is a posynomial function, the SINR constraint in (26b) cannot be transformed into the format as that in (36b), in which the right hand side is a monomial function. Hence, the problem cannot be directly transformed into a GP problem. To solve this issue, we introduce Theorem 3 as follows.

Theorem 3: For any given vector $\breve{\mathbf{x}} = \{\breve{x}_1, \cdots, \breve{x}_K\}$ with $\breve{x}_i \geq 0, \forall n$, function $W(\mathbf{x}) = \prod_{i\in\mathcal{K}}\left(1 + x_i\right)$ is lower bounded by

$$W(\mathbf{x}) = \prod_{i\in\mathcal{K}}\left(1 + x_i\right) \geq \lambda \prod_{i\in\mathcal{K}} x_i^{\tau_i} \triangleq Y(\mathbf{x}) \qquad (42)$$

where λ and $\tau_i, \forall i$ are given by

$$\lambda = \frac{\prod_{i\in\mathcal{K}}\left(1 + \breve{x}_i\right)}{\prod_{i\in\mathcal{K}} \breve{x}_i^{\tau_i}}, \tau_i = \frac{\breve{x}_i}{1 + \breve{x}_i}, \forall i, \qquad (43)$$

and $\mathbf{x}$ is given by $\mathbf{x} = \{x_1, \cdots, x_K\}$.

In addition, we have:

$$W(\breve{\mathbf{x}}) = Y(\breve{\mathbf{x}}), \nabla W(\breve{\mathbf{x}}) = \nabla Y(\breve{\mathbf{x}}). \qquad (44)$$

where $\nabla W(\mathbf{x})$ and $\nabla Y(\mathbf{x})$ denote the gradient of function $W(\cdot)$ and $Y(\cdot)$ w.r.t. $\mathbf{x}$, respectively.

Proof: Please refer to Appendix E. ■

Based on Theorem 3, we replace the polynomial functions in the nominator of $\hat{\gamma}_k$ in (41) with their best local monomial approximations by employing Theorem 3 [2]. In specific, we denote $\mathbf{P}^{(n)} = \{p_k^{p(n)}, p_k^{d(n)}, \forall k\}$ as the power allocation in the n-th iteration, and the corresponding χ_k is given by $\chi_k^{(n)}$. Then, in the $n+1$-th iteration, we approximate the term $\prod_{i\in\mathcal{K}} (1 + \alpha_i K p_i^p)$ in the nominator of (41) by its best local monomial approximations, which is given by function $Y(\mathbf{x})$ in Theorem 3 with $x_i = \alpha_i K p_i^p$:

$$\prod_{i\in\mathcal{K}} (1 + \alpha_i K p_i^p) \geq \lambda^{(n)} \prod_{i\in\mathcal{K}} (\alpha_i K p_i^p)^{\tau_i^{(n)}}, \tag{45}$$

where $\lambda^{(n)}$ and $\tau_i^{(n)}, \forall i$ are given in (43) with $\tilde{x}_i = \alpha_i K p_i^{p(n)}, \forall i$. Then, we focus on the following constraint instead of the original SINR constraint in (26b):

$$\chi_k \left(1 + \alpha_k K p_k^p\right) \left(\sum_{i\in\mathcal{K}} p_i^d \alpha_i \prod_{j\neq i} (1 + \alpha_j K p_j^p) + \prod_{i\in\mathcal{K}} (1 + \alpha_i K p_i^p)\right) \tag{46}$$

$$\leq (M - K)\alpha_k^2 K p_k^p p_k^d \lambda^{(n)} \prod_{i\in\mathcal{K}} (\alpha_i K p_i^p)^{\tau_i^{(n)}}. \tag{47}$$

Note that the left hand side (LHS) of (46) is a posynomial function, and the right hand side (RHS) becomes a monomial function. In addition, the LHS is no larger than RHS. Hence, constraint (46) satisfies the conditions for a problem to be a GP problem [26].

Then, by using the same method as in the MRC case to deal with the OF, in the $n+1$-th iteration, we aim to solve the following GP problem:

$$\max_{\{\chi_k\},\{p_k^p\},\{p_k^d\}} \prod_{k\in\mathcal{K}} \chi_k^{\hat{w}_k^{(n)}} \tag{48a}$$

$$\text{s.t.} \quad (24c), (46) \tag{48b}$$

$$\hat{\gamma}_k \geq 1 \Big/ f_k^{-1}\left(\frac{R_k^{\text{req}} \ln 2}{1 - \beta}\right), \forall k, \tag{48c}$$

where the parameters $\hat{w}_k^{(n)}$'s are the same as those in the MRC case. This problem can be efficiently solved by using CVX [27].

Based on the above discussion, the iterative algorithm to solve Problem (26) for the case of ZF is provided in Algorithm 2.

---

[2]The best local monomial approximations means that the approximation function should satisfy three conditions as specified in Section IV-A of [32].

Algorithm 2 Iterative algorithm for solving Problem (26) for ZF

1: Initialize iteration number $n = 1$, error tolerance ξ. Initialize a feasible power allocation $\{p_k^{p(0)}, p_k^{d(0)}, \forall k\}$, calculate $\{\chi_k^{(0)}, \rho_1^{(0)}, \rho_2^{(0)}, \hat{w}_k^{(0)}, \lambda^{(0)}, \tau_k^{(0)}, \forall k\}$, and calculate the OF of Problem (26), denoted as $\text{Obj}^{(0)}$.

2: With given $\{\chi_k^{(n-1)}, \rho_1^{(n-1)}, \rho_2^{(n-1)}, \hat{w}_k^{(n-1)}, \lambda^{(n-1)}, \tau_k^{(n-1)} \forall k\}$, solve Problem (48) by using the CVX package to obtain $\{p_k^{p(n)}, p_k^{d(n)}, \chi_k^{(n)}, \forall k\}$.

3: Update $\{\rho_1^{(n)}, \rho_2^{(n)}, \hat{w}_k^{(n)}, \lambda^{(n)}, \tau_k^{(n)}, \forall k\}$;

4: Calculate new OF $\text{Obj}^{(n)}$. If $|\text{Obj}^{(n)} - \text{Obj}^{(n-1)}|/\text{Obj}^{(n)} < \xi$, terminate. Otherwise, set $n \leftarrow n + 1$, go to step 2.

B. Algorithm Analysis

Algorithm 2 can be analyzed similar to Algorithm 1 except the convergence analysis since the problems to be solved in each iteration of Algorithm 2 do not have the same set of constraints.

In the following, we prove that the solution obtained in the n-th iteration is also feasible for the problem to be solved in the $n + 1$-th iteration. We only need to check constraint (46) since the other two constraints are the same in each iteration.

Let us denote $\{\chi_k^{(n)}, p_k^{p(n)}, p_k^{d(n)}, \forall k\}$ as the optimal solution in the n-th iteration. Then, it is also feasible solution, and we have

$$\chi_k^{(n)} \left(1 + \alpha_k K p_k^{p(n)}\right) \left(\sum_{i \in \mathcal{K}} p_i^{d(n)} \alpha_i \prod_{j \neq i} (1 + \alpha_j K p_j^{p(n)}) + \prod_{i \in \mathcal{K}} (1 + \alpha_i K p_i^{p(n)})\right)$$

$$\leq (M - K) \alpha_k^2 K p_k^{p(n)} p_k^{d(n)} \lambda^{(n-1)} \prod_{i \in \mathcal{K}} \left(\alpha_i K p_i^{p(n)}\right)^{\tau_i^{(n-1)}}. \tag{49}$$

By using (45), we have

$$\prod_{i \in \mathcal{K}} (1 + \alpha_i K p_i^{p(n)}) \geq \lambda^{(n-1)} \prod_{i \in \mathcal{K}} \left(\alpha_i K p_i^{p(n)}\right)^{\tau_i^{(n-1)}}. \tag{50}$$

In addition, by using (44) in Theorem 3 and (50), we have

$$\lambda^{(n)} \prod_{i \in \mathcal{K}} \left(\alpha_i K p_i^{p(n)}\right)^{\tau_i^{(n)}} = \prod_{i \in \mathcal{K}} (1 + \alpha_i K p_i^{p(n)}) \geq \lambda^{(n-1)} \prod_{i \in \mathcal{K}} \left(\alpha_i K p_i^{p(n)}\right)^{\tau_i^{(n-1)}}. \tag{51}$$

Finally, by combining (46) and (51), we have

$$\chi_k^{(n)} \left(1 + \alpha_k K p_k^{p(n)}\right) \left(\sum_{i \in \mathcal{K}} p_i^{d(n)} \alpha_i \prod_{j \neq i} (1 + \alpha_j K p_j^{p(n)}) + \prod_{i \in \mathcal{K}} (1 + \alpha_i K p_i^{p(n)})\right)$$

$$\leq (M - K) \alpha_k^2 K p_k^{p(n)} p_k^{d(n)} \lambda^{(n)} \prod_{i \in \mathcal{K}} \left(\alpha_i K p_i^{p(n)}\right)^{\tau_i^{(n)}}. \tag{52}$$

Hence, $\{\chi_k^{(n)}, p_k^{p(n)}, p_k^{d(n)}, \forall k\}$ is also a feasible solution in the $n + 1$-th iteration. Then, by using the similar proof as in the case of MRC, we can also prove that Algorithm 2 is guaranteed

to converge. By using a similar method, we also prove that this algorithm will converge to a feasible solution of Problem (26).

V. SIMULATION RESULTS

In this section, we provide simulation results to demonstrate the effectiveness of our proposed algorithms for industrial automation systems. The channel path loss is modeled as $PL = 35.3 + 37.6\log_{10}d$ (dB) [33], and the small-scale fading is modeled as Rayleigh fading with zero mean and unit variance. Unless otherwise specified, the simulation parameters are set as follows: number of transmit antennas of $M = 100$, number of devices of $K = 10$, channel bandwidth of $B = 0.2$ MHz, noise power spectral density of -174 dBm/Hz, decoding error probability of $\varepsilon_k = 10^{-9}, \forall k$, number of transmit antennas of $M = 100$, and maximum transmission duration of 0.5 ms. The maximum blocklength is calculated as $L = BT = 100$. The other parameters are specified in each figure. The energy constraint for each device is assumed to be equal, i.e., $E_k = E, \forall k$, and each device has the same data rate targets, $R_k^{\text{req}} = R^{\text{req}}, \forall k$. The weights for each device are uniformly generated within $[0, 1]$.

A. Tightness of the Date Rate LB

In Fig. 5 and Fig. 6, we investigates the tightness of the LB derived for the cases of MRC and ZF, respectively. The simulation results are obtained through the Monte-Carlo simulation by averaging over 5000 random channel generations. It is observed that the data rate LB is tight for both cases of MRC and ZF for any number of transmit antennas. Interestingly, the curves for the ZF case are almost overlapped with each other. This verifies that the data rate LBs derived in Theorem 1 and Theorem 2 are suitable for optimization instead of directly optimizing the complicated expectation expression.

B. Convergence Behaviour of the Proposed Algorithms

Fig. 7 and Fig. 8 investigates the impact of the energy limit at each device on the performance of the proposed Algorithm 1 for the MRC case and Algorithm 2 for the ZF case, respectively. From these two figures, we can observe that both algorithms converge rapidly for various energy limits and only 2 or 3 iterations are sufficient for the algorithms to converge. This demonstrates the low-complexity of the proposed algorithms.
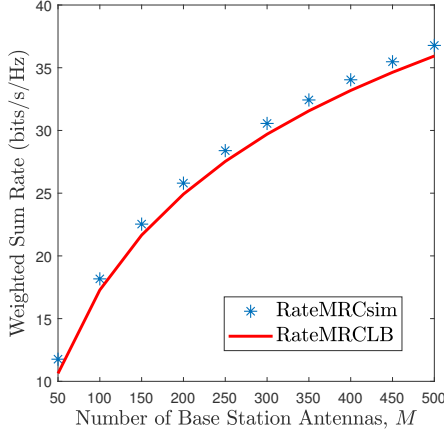
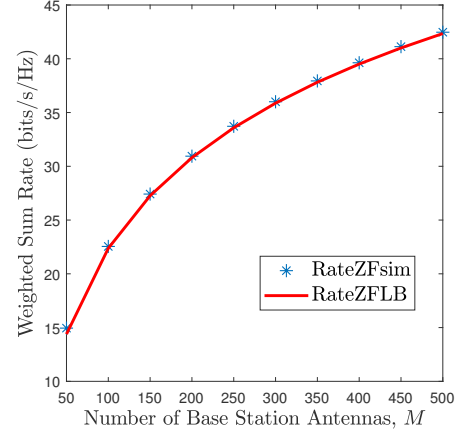Fig. 5: Tightness of the derived data rate LB for the MRC case.

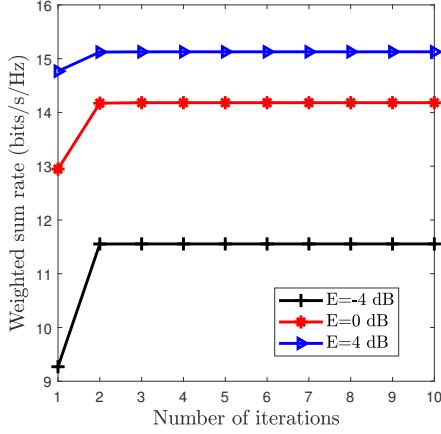Fig. 6: Tightness of the derived data rate LB for the ZF case.

Fig. 7: Convergence behaviour of the proposed algorithm for the MRC case.
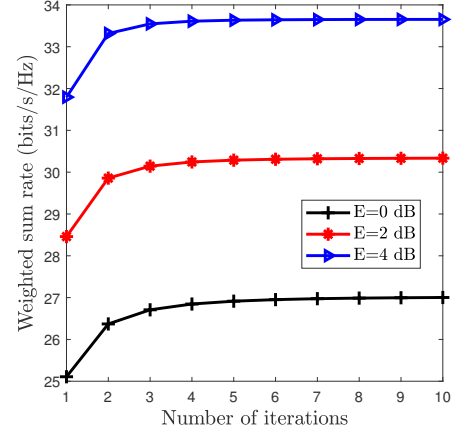

Fig. 8: Convergence behaviour of the proposed algorithm for the ZF case.

C. Performance Comparison

In this subsection, we compare the proposed algorithms with the following algorithms:

- **Upper bound:** In this method, Shannon capacity is adopted for optimization in Problem (26). In other words, the penalty term $a_k G(\chi_k)$ is set to zero in both the OF and rate constraint (24b), i.e., $a_k G(\chi_k) = 0, \forall k$. This method provides the upper bound of the average weighted sum rate performance in the IIoT networks.

- **Conventional alg.**: As in [34], the solution obtained from the upper bound is applied in (20) and (22) for the cases of MRC and ZF respectively by considering the penalty term $a_k G(\chi_k)$. That means that the upper bound is used for obtaining solutions, but the achievable

data rate under finite blocklength is used for performance evaluation.

- **Fixed pilot power alg.**: In this scheme, we only optimize the data transmission power p_k^d while fixing the pilot power as $p_k^p = E/L$. This algorithm is provided to show the benefits of jointly optimizing the transmission and pilot power.

The following results are obtained by averaging over 100 Monte-Carlo simulations where in each snapshot the devices are randomly generated in the cell. For each snapshot, if the device's achievable data rate cannot achieve its rate targets, we set the corresponding data rate to zero. The proposed algorithm is denoted as 'Proposed alg.' in the following figures.

In Fig. 9 and Fig. 10, we show the average weighted sum rate versus the energy limit at each device, E. The rate targets for MRC and ZF are set as $R^{\mathrm{req}} = 1$ bits/s/Hz and $R^{\mathrm{req}} = 4$ bits/s/Hz, respectively. As seen from these figures, the system performance increases with the available energy at each device since the SINR at each device is increased. Note that the weighted sum rate achieved by some algorithms may approach zero, which means the power allocation solution is infeasible. As expected, the upper bound has the best performance since the penalty terms are not considered. Furthermore, the 'conventional alg.' based on Shannon capacity formula has higher probability to violate the data rate requirement, especially for samll E. Therefore, Shannon capacity cannot be employed for the transmission design of URLLC for industrial applications, in particular when the energy limit is small as the QoS target cannot be guaranteed. However, for the large value of E, the SINR value for each device is very high and then the penalty term $a_k G(\chi_k)$ can be ignored when comparing with the first term of $\ln(1 + \chi_k)$. Hence, the 'conventional alg.' will achieve a similar performance as that of the proposed algorithm. As shown in Fig. 9 and Fig. 10, by jointly optimizing the data transmission power and pilot power, the proposed algorithm is superior over the 'Fixed pilot power alg.', which only optimizes the data transmission power. The performance gain is obvious when the energy limit is low, especially for the case of ZF. This means the need of jointly optimizing pilot and data power at low energy limit. This can be explained as follows. When the energy limit is low, the system performance is limited by channel estimation procedure. Through the joint power allocation, some power can be borrowed from that for data transmission to enhance the channel estimation accuracy, and thus increases the weighted sum rate. However, in the high energy E, the accuracy of channel estimation is already enough, and additional joint power control bring marginal performance gain. Another interesting observation is that the performance gain of the proposed algorithm
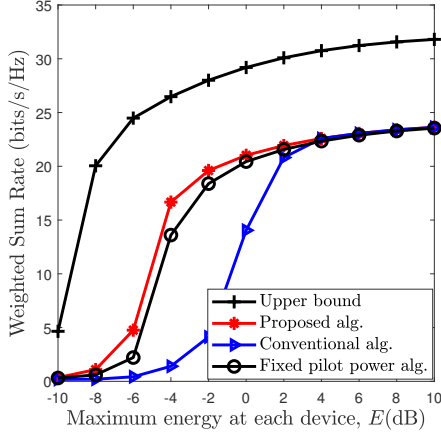
Fig. 9: Average weighted sum rate vs. maximum energy limit $10\log_{10}E$ for various schemes for the case of MRC.
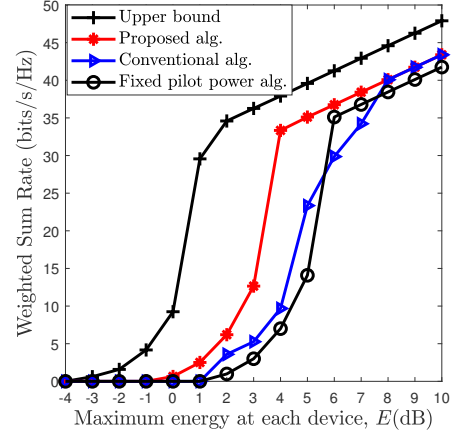

Fig. 10: Average weighted sum rate vs. maximum energy limit $10\log_{10}E$ for various schemes for the case of ZF.
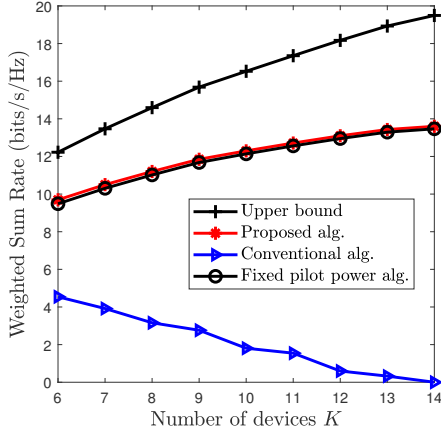

Fig. 11: Average weighted sum rate vs. the number of devices for various schemes for the case of MRC.
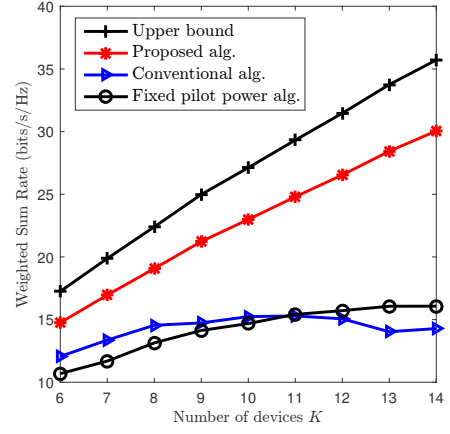

Fig. 12: Average weighted sum rate vs. the number of devices for various schemes for the case of ZF.

over 'Fixed pilot power alg.' is much smaller for the MRC case than the ZF case. This may be due to the fact that the CSI is more important when using it for removing multiple device interference in the ZF case.

Fig. 11 and Fig. 12 show the average weighted sum rate versus the number of devices for various schemes. The rate targets for MRC and ZF are set as $R^{\text{req}} = 1$ bits/s/Hz and $R^{\text{req}} = 2$ bits/s/Hz, respectively. The energy limit for MRC and ZF are set as $E = 2$ and $E = 1$, respectively. It is observed from Fig. 11 and Fig. 12 that the average weighted sum rate achieved

by all the algorithms except the 'Conventional alg.' increases with the number of devices since these schemes fully exploit the multi-device diversity. For the MRC case, the performance of the 'Conventional alg.' decreases with the number of devices. The main reason is that the design is based on Shannon capacity formula, which does not take into consideration the effect of short blocklength on the achievable data rate. As a result, the probability that the achieved data rate violates the rate requirement will increase with the number of devices. It is interesting to observe that 'Fixed pilot power alg.' has similar performance as the proposed algorithm for the case of MF. However, for the case of ZF, the proposed algorithm significantly outperforms 'Fixed pilot power alg.', and the performance gain is increasing with the number of devices. This again reveals the importance of joint power optimization in the case of ZF. For the ZF case, the weighted sum rate achieved by 'Conventional alg.' is first increasing with the number of devices and then decreases with it.

Finally, we study the effect of blocklength on the weighted sum rate performance in Fig. 13 and Fig. 14 for the cases of MRC and ZF, respectively. The rate targets for MRC and ZF are set as $R^{\mathrm{req}} = 2$ bits/s/Hz and $R^{\mathrm{req}} = 4$ bits/s/Hz, respectively. As expected, the system performance increases with the blocklength since more time/frequency resource can be exploited for transmission. Some interesting observations can be found in these figures. First, when the blocklength is small, there is significant performance gap between the proposed algorithm and the upper bound. However, this gap starts to reduce with the increase of the blocklength. This demonstrates that the blocklength has significant impact for IIoT devices with URLLC requirements. In addition, the proposed algorithm outperforms the 'Fixed pilot power alg.' algorithm for both cases of MRC and ZF. This may be due to the fact that higher rate target is imposed in this example than that in Fig. 11.

VI. Conclusions

In this paper, we studied the resource allocation for uplink massive MIMO systems to support critical IIoT operating under finite channel blocklength, where multiple robots and/or actuators transmit URLLC signals to the central controller simultaneously. We first derived the closed-form data rate LB with imperfect CSI for both MRC and ZF receivers under the short packet transmission. We then formulate the weighted sum rate maximization to jointly optimize the pilot and payload power allocation while considering their energy, minimum data rate and decoding
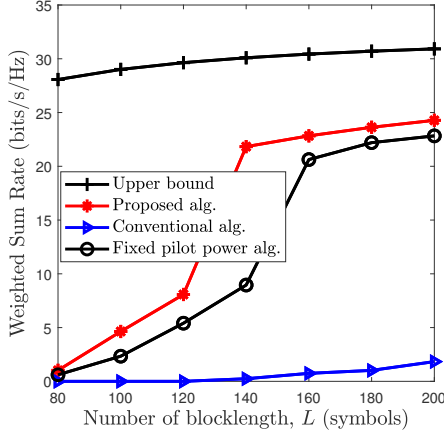
Fig. 13: Average weighted sum rate vs. block-length for various schemes for the case of MRC.
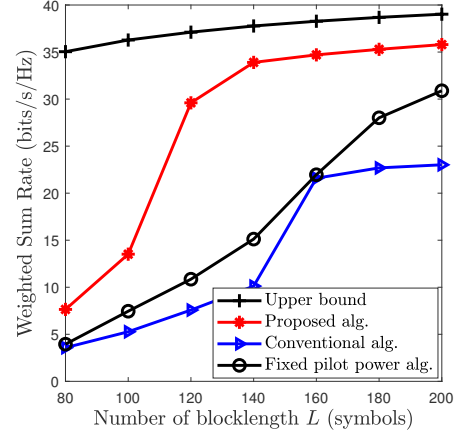
Fig. 14: Average weighted sum rate vs. block-length for various schemes for the case of ZF.

error probability requirements. This optimization problem is non-convex, and we proposed novel low-complexity iterative algorithms to solve it. Simulation results demonstrate that the algorithm converges rapidly, and outperforms the existing benchmark algorithms, especially the algorithm based on conventional Shannon capacity. This reveals the importance of adopting the achievable data expression for finite channel blocklength.

APPENDIX A

PROOF OF THEOREM 1

We first consider the MRC beamforming. The proof follows the similar steps as those in Appendix A of [35] for perfect CSI. Denote γ_k as the instantaneous SINR value when using MRC. By substituting $\mathbf{a}_k = \hat{\mathbf{h}}_k$ into (12), we have

$$\gamma_k = \frac{p_k^d \left\| \hat{\mathbf{h}}_k \right\|^4}{\sum_{i \in \mathcal{K} \backslash k} p_i^d \left| \hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_i \right|^2 + \sum_{i \in \mathcal{K}} p_i^d \left| \hat{\mathbf{h}}_k^H \tilde{\mathbf{h}}_i \right|^2 + \left\| \hat{\mathbf{h}}_k \right\|^2}. \tag{53}$$

Then, $\mathbb{E}\left\{ \frac{1}{\gamma_k} \right\}$ can be expressed as

$$\mathbb{E}\left\{ \frac{1}{\gamma_k} \right\} = \mathbb{E}\left\{ \frac{\sum_{i \in \mathcal{K} \backslash k} p_i^d |u_{k,i}|^2 + \sum_{i \in \mathcal{K}} p_i^d |v_{k,i}|^2 + 1}{p_k^d \left\| \hat{\mathbf{h}}_k \right\|^2} \right\} \tag{54}$$

where $u_{k,i} = \frac{\hat{\mathbf{h}}_k^H \hat{\mathbf{h}}_i}{\left\| \hat{\mathbf{h}}_k \right\|}$ and $v_{k,i} = \frac{\hat{\mathbf{h}}_k^H \tilde{\mathbf{h}}_i}{\left\| \hat{\mathbf{h}}_k \right\|}$. Conditioned on $\hat{\mathbf{h}}_k$, $u_{k,i}$ and $v_{k,i}$ are Gaussian random variables with zero mean and variance equal to σ_i and δ_i, respectively. In addition, $u_{k,i}$ and $v_{k,i}$

are independent of $\hat{\mathbf{h}}_k$. Then, we have

$$\mathbb{E}\left\{\frac{1}{\gamma_k}\right\} = \left(\sum_{i\in\mathcal{K}\backslash k} p_i^d\sigma_i + \sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\mathbb{E}\left\{\frac{1}{p_k^d\left\|\hat{\mathbf{h}}_k\right\|^2}\right\}. \tag{55}$$

By using the identity [36]

$$\mathbb{E}\left\{\mathrm{tr}\left(\mathbf{W}^{-1}\right)\right\} = \frac{m}{n-m} \tag{56}$$

where $\mathbf{W} \sim \mathcal{W}_m\left(n, \mathbf{I}_n\right)$ is an $m \times m$ central complex Wishart matrix with n $(n-m)$ degrees of freedom. Then, we have

$$\mathbb{E}\left\{\frac{1}{p_k^d\left\|\hat{\mathbf{h}}_k\right\|^2}\right\} = \frac{1}{p_k^d(M-1)\sigma_k}, \text{ for } M \geq 2. \tag{57}$$

By substituting (57) into (55), we can obtain $\hat{\gamma}_k$ in (21).

APPENDIX B

PROOF OF THEOREM 2

By using ZF, we have $\mathbf{A}^H\hat{\mathbf{H}} = \mathbf{I}_K$. Thus, $\mathbf{a}_k\hat{\mathbf{h}}_i$ is equal to one if $k = i$; otherwise, it is equal to zero. Then, the instantaneous SINR for ZF can be rewritten as

$$\gamma_k = \frac{p_k^d}{\sum_{i\in\mathcal{K}} p_i^d\left|\mathbf{a}_k^H\tilde{\mathbf{h}}_i\right|^2 + \|\mathbf{a}_k\|^2}. \tag{58}$$

Define $\mathbf{\Lambda} = \mathrm{diag}\left\{\sqrt{\sigma_1}, \cdots, \sqrt{\sigma_K}\right\}$ and $\breve{\mathbf{H}} = \left[\frac{\hat{\mathbf{h}}_1}{\sqrt{\sigma_1}}, \cdots, \frac{\hat{\mathbf{h}}_K}{\sqrt{\sigma_K}}\right]$. Then, we can express $\hat{\mathbf{H}}$ as $\hat{\mathbf{H}} = \breve{\mathbf{H}}\mathbf{\Lambda}$. The columns of $\breve{\mathbf{H}}$ are independent of each other, and each column follows the distribution of $\mathcal{CN}(\mathbf{0}, \mathbf{I})$. Then, $\mathbb{E}\left\{\frac{1}{\gamma_k}\right\}$ can be expressed as

$$\mathbb{E}\left\{\frac{1}{\gamma_k}\right\} = \left(\sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\mathbb{E}\left\{\left[\left(\hat{\mathbf{H}}^H\hat{\mathbf{H}}\right)^{-1}\right]_{k,k}\right\}\bigg/ p_k^d \tag{59}$$

$$= \left(\sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\mathbb{E}\left\{\left[\left(\mathbf{\Lambda}\breve{\mathbf{H}}^H\breve{\mathbf{H}}\mathbf{\Lambda}\right)^{-1}\right]_{k,k}\right\}\bigg/ p_k^d \tag{60}$$

$$= \left(\sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\mathbb{E}\left\{\left[\left(\breve{\mathbf{H}}^H\breve{\mathbf{H}}\right)^{-1}\right]_{k,k}\right\}\bigg/ \sigma_k p_k^d \tag{61}$$

$$= \left(\sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\mathbb{E}\left\{\mathrm{tr}\left[\left(\breve{\mathbf{H}}^H\breve{\mathbf{H}}\right)^{-1}\right]\right\}\bigg/ K\sigma_k p_k^d \tag{62}$$

$$= \left(\sum_{i\in\mathcal{K}} p_i^d\delta_i + 1\right)\bigg/(M-K)\sigma_k p_k^d \tag{63}$$

where (63) is obtained by using (56).

APPENDIX C

PROOF OF LEMMA 2

For notational simplicity, we define function $T(x) \triangleq 1 - \frac{1}{(1+x)^2}$. The second derivative of $T(x)$ w.r.t. x is given by

$$T''(x) = \frac{-6}{(x+1)^4} \leq 0. \tag{64}$$

Next, we prove that $G(x)$ is a concave function of x. Since $T(x)$ is a concave function w.r.t. x, we have

$$\theta T(\hat{x}) + (1 - \theta)T(\tilde{x}) \leq T(\theta\hat{x} + (1 - \theta)\tilde{x}), \tag{65}$$

for $0 \leq \theta \leq 1$, where $\hat{x}$ and $\tilde{x}$ are two different non-negative values. On the other hand, $\sqrt{x}$ is a concave function w.r.t. x. Then, for $0 \leq \theta \leq 1$ we have

$$\theta\sqrt{T(\hat{x})} + (1 - \theta)\sqrt{T(\tilde{x})} \leq \sqrt{\theta T(\hat{x}) + (1 - \theta)T(\tilde{x})}. \tag{66}$$

By combining (66) with (65), we have

$$\theta\sqrt{T(\hat{x})} + (1 - \theta)\sqrt{T(\tilde{x})} \leq \sqrt{T(\theta\hat{x} + (1 - \theta)\tilde{x})}, \tag{67}$$

which is equivalent to $\theta G(\hat{x}) + (1 - \theta)G(\tilde{x}) \leq G(\theta\hat{x} + (1 - \theta)\tilde{x})$. Hence, $G(x)$ is a concave function w.r.t. x, which completes the proof.

APPENDIX D

PROOF OF LEMMA 3

The equalities in (30) can be readily proved by substituting the expressions of ρ and η in (28) and (29) into (27). Next, we focus on the proof of Inequality (27).

Define function $H(x) \triangleq F(x) - G(x)$, and thus $H(\tilde{x}) = 0$. Then, with some simple manipulations, the first-order derivative of function $H(x)$ w.r.t. x can be calculated as

$$H'(x) = \frac{\tilde{x}\sqrt{x^2 + 2x}(x+1)^2 - x\sqrt{\tilde{x}^2 + 2\tilde{x}}(\tilde{x}+1)^2}{\sqrt{\tilde{x}^2 + 2\tilde{x}}(\tilde{x}+1)^2 x\sqrt{x^2 + 2x}(x+1)^2}. \tag{68}$$

Since both x and $\tilde{x}$ are positive values, the sign of $H'(x)$ only depends on the nominator of $H'(x)$. Then, denote the nominator of $H'(x)$ as $J(x)$.

The function $J(x)$ can be rewritten as

$$J(x) = x\tilde{x}\left(\sqrt{1 + \frac{2}{x}}(x+1)^2 - \sqrt{1 + \frac{2}{\tilde{x}}}(\tilde{x}+1)^2\right). \tag{69}$$

Next, we show that $J(x) \leq 0$ when $\frac{\sqrt{17}-3}{4} \leq x \leq \tilde{x}$, and $J(x) > 0$ when $x > \tilde{x}$. In specific, let us define

$$U(x) \triangleq \sqrt{1 + \frac{2}{x}(x+1)^2}. \tag{70}$$

The first-order derivative of $U(x)$ w.r.t. x is given by

$$U'(x) = \frac{x+1}{x^{1.5}\sqrt{x+2}}\left(2x^2 + 3x - 1\right). \tag{71}$$

It can be readily proved that $U'(x) \leq 0$ when $0 < x \leq \frac{\sqrt{17}-3}{4}$, and $U'(x) > 0$ when $x > \frac{\sqrt{17}-3}{4}$.

As a result, when $\frac{\sqrt{17}-3}{4} \leq x \leq \tilde{x}$, we have $U(x) < U(\tilde{x})$, and thus $J(x) < 0$. From (68), we know that $H'(x) < 0$, which means $H(x)$ is a monotonically decreasing function of x. Hence, $H(x) \geq H(0) = 0$ holds for $\frac{\sqrt{17}-3}{4} \leq x \leq \tilde{x}$, which equivalently means $F(x) \geq G(x)$. On the other hand, when $x > \tilde{x}$, we have $U(x) > U(\tilde{x})$, and thus $J(x) > 0$, which means $H'(x) > 0$ and $H(x)$ is a monotonically increasing function of x. Hence, $H(x) > H(0) = 0$ holds for $x \geq \tilde{x}$, and thus $F(x) > G(x)$.

Based on the above analysis, when $x \geq \frac{\sqrt{17}-3}{4}$, $F(x)$ is always no smaller than $G(x)$, which completes the proof.

APPENDIX E

PROOF OF THEOREM 3

The first equation in (44) can be readily verified. Then, we focus on the second equality in (44). The partial derivative of $W(\check{\mathbf{x}})$ and $Y(\check{\mathbf{x}})$ w.r.t. x_i are given by

$$\frac{\partial W(\mathbf{x})}{\partial x_i} = \prod_{j \neq i}(1 + x_j), \frac{\partial Y(\mathbf{x})}{\partial x_i} = \lambda \tau_i x_i^{-1} \prod_{j \in \mathcal{K}} x_j^{\tau_j}, i = 1, \cdots, K. \tag{72}$$

By substituting $\mathbf{x} = \check{\mathbf{x}}$, λ and $\tau_i, \forall i$ into into the above two functions, we can show that

$$\left.\frac{\partial W(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\check{\mathbf{x}}} = \left.\frac{\partial Y(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\check{\mathbf{x}}}, i = 1, \cdots, K. \tag{73}$$

Hence, the second equality in (44) is proved.

Now, we begin to prove (42). Before proceeding, we first introduce the following lemma.

Lemma 5: For any given $\check{x} \geq 0$, we have the following inequality:

$$\frac{1+x}{x^\tau} \geq \frac{1+\check{x}}{\check{x}^\tau}, \tag{74}$$

where τ is given by $\tau = \frac{\check{x}}{1+\check{x}}$, and equality holds only when $x = \check{x}$.

Proof: The proof is similar to those in Lemma 3, and thus omitted for simplicity. ∎

By applying inequality (74) for each $x_i, i = 1, \cdots, K$, we have

$$\frac{1 + x_i}{x_i^{\tau_i}} \geq \frac{1 + \breve{x}_i}{x_i^{\tau_i}}, i = 1, \cdots, K. \tag{75}$$

Then, by multiplying the above K inequalities, we have

$$\prod_{i \in \mathcal{K}} \frac{1 + x_i}{x_i^{\tau_i}} \geq \prod_{i \in \mathcal{K}} \frac{1 + x_i}{x_i^{\tau_i}} \tag{76}$$

which completes the proof.

References

[1] R. Drath and A. Horch, "Industrie 4.0: Hit or hype?[industry forum]," *IEEE industrial electronics mag.*, vol. 8, no. 2, pp. 56–58, 2014.

[2] Z. Pang, M. Luvisotto, and D. Dzung, "Wireless high-performance communications: The challenges and opportunities of a new target," *IEEE Industrial Electronics Mag.*, vol. 11, no. 3, pp. 20–25, 2017.

[3] M. Luvisotto, Z. Pang, and D. Dzung, "Ultra high performance wireless control for critical applications: Challenges and directions," *IEEE Trans. Industrial Informatics*, vol. 13, no. 3, pp. 1448–1459, 2016.

[4] Z. Ma, M. Xiao, Y. Xiao, Z. Pang, H. V. Poor, and B. Vucetic, "High-reliability and low-latency wireless communication for internet of things: Challenges, fundamentals and enabling technologies," *IEEE Internet Things J.*, 2019.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdu, "Channel Coding Rate in the Finite Blocklength Regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[6] C. She, C. Yang, and T. Q. Quek, "Radio resource management for ultra-reliable and low-latency communications," *IEEE Commun. Mag.*, vol. 55, no. 6, pp. 72–78, 2017.

[7] X. Sun *et al.*, "Short-Packet Downlink Transmission With Non-Orthogonal Multiple Access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, July 2018.

[8] Y. Hu *et al.*, "SWIPT-Enabled Relaying in IoT Networks Operating with Finite Blocklength Codes," *IEEE J. Sel. Areas in Commun.*, pp. 1–1, 2018.

[9] 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, March.

[10] H. Ren, C. Pan, Y. Deng, M. Elkashlan, and A. Nallanathan, "Resource Allocation for Ultra-Reliable and Low-Latency Communications in 5G Mission-Critical IoT Networks," in *IEEE ICC*, 2019.

[11] J. Chen, L. Zhang, Y.-C. Liang, X. Kang, and R. Zhang, "Resource allocation for wireless-powered IoT networks with short packet communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 2, pp. 1447–1461, 2019.

[12] C. She, C. Yang, and T. Q. S. Quek, "Joint Uplink and Downlink Resource Configuration for Ultra-Reliable and Low-Latency Communications," *IEEE Trans. Commun.*, vol. 66, no. 5, pp. 2266–2280, May 2018.

[13] W. R. Ghanem, V. Jamali, Y. Sun, and R. Schober, "Resource allocation for multi-user downlink URLLC-OFDMA systems," *arXiv preprint arXiv:1901.05825*, 2019.

[14] 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 Industrial Electronics Mag.*, vol. 12, no. 2, pp. 50–58, 2018.

[15] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, "What will 5G be?" *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, 2014.

[16] H. V. Cheng, E. Björnson, and E. G. Larsson, "Optimal pilot and payload power control in single-cell massive MIMO systems," *IEEE Trans. Signal Process.*, vol. 65, no. 9, pp. 2363–2378, 2017.

[17] O. Saatlou, M. O. Ahmad, and M. Swamy, "Joint data and pilot power allocation for massive mu-mimo downlink tdd systems," *IEEE Trans. Circuits and Systems II: Express Briefs*, vol. 66, no. 3, pp. 512–516, 2019.

[18] S. Lu and Z. Wang, "Joint optimization of power allocation and training duration for uplink multiuser mimo communications," in *2015 IEEE WCNC*. IEEE, 2015, pp. 322–327.

[19] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Communs.*, vol. 9, no. 11, pp. 3590–3600, November 2010.

[20] J. Shen, J. Zhang, and K. B. Letaief, "Downlink user capacity of massive MIMO under pilot contamination," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3183–3193, June 2015.

[21] M. Mousaei and B. Smida, "Optimizing pilot overhead for ultra-reliable short-packet transmission," in *IEEE ICC*, 2017, pp. 1–5.

[22] 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. Letter,*, 2019.

[23] Z.-Q. Luo and S. Zhang, "Dynamic spectrum management: Complexity and duality," *IEEE J. Sel. Topics Signal Process.*, vol. 2, no. 1, pp. 57–73, 2008.

[24] H. V. Cheng, E. Björnson, and E. G. Larsson, "Uplink pilot and data power control for single cell massive MIMO systems with MRC," in *IEEE ISWCS*, 2015, pp. 396–400.

[25] O. N. Yilmaz, Y.-P. E. Wang, N. A. Johansson, N. Brahmi, S. A. Ashraf, and J. Sachs, "Analysis of ultra-reliable and low-latency 5G communication for a factory automation use case," in *IEEE ICCW*, 2015, pp. 1190–1195.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[27] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.1," 2014.

[28] E. Matskani, N. D. Sidiropoulos, Z.-q. Luo, and L. Tassiulas, "Convex approximation techniques for joint multiuser downlink beamforming and admission control," *IEEE Trans. Wireless Commun.*, vol. 7, no. 7, pp. 2682–2693, 2008.

[29] 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, 2017.

[30] S. Boyd, S.-J. Kim, L. Vandenberghe, and A. Hassibi, "A tutorial on geometric programming," *Optimization and engineering*, vol. 8, no. 1, p. 67, 2007.

[31] S. Shi, M. Schubert, and H. Boche, "Rate optimization for multiuser MIMO systems with linear processing," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 4020–4030, 2008.

[32] M. Chiang, C. W. Tan, D. P. Palomar, D. ONeill, and D. Julian, "Power control by geometric programming," *Resource Allocation in Next Generation Wireless Networks*, vol. 5, pp. 289–313, 2005.

[33] E. U. T. R. Access, "Further advancements for E-UTRA physical layer aspects," *3GPP TR 36.814, Tech. Rep.*, 2010.

[34] G. W. R, S. Y. Jamali V, and R. Schober, "Resource allocation for multi-user downlink URLLC-OFDMA systems," *arXiv preprint arXiv:1901.05825, 2019*, 2016.

[35] 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, 2013.

[36] A. M. Tulino, S. Verdú *et al.*, "Random matrix theory and wireless communications," *Foundations and Trends® in Commun. and Inf. Theory*, vol. 1, no. 1, pp. 1–182, 2004.