Achieving mURLLC Under Nakagami-m Fading in the CF mMIMO System

Qingqin Xu[1], Zhuoer Li[2], Biru Zhang[1], and Jie Zeng[2,3](✉)

[1] School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China
{s210131272,s220132211}@stu.cqupt.edu.cn

[2] School of Cyberspace Science and Technology, Beijing Institute of Technology, Beijing 100081, China
1120203294@bit.edu.cn

[3] Department of Electronic Engineering, Tsinghua University, Beijing 100084, China
zengjie@tsinghua.edu.cn

Abstract. Massive ultra-reliable and low latency communications (mURLLC) is emerging as a new important and critical service category in the six generation (6G) mobile communications. However, it is extremely challenging to achieve mURLLC because it needs to fulfill three performance requirements simultaneously, i.e., reliability, latency, and large-scale access requirements. Existing related studies are few and almost all based on the ideal assumption of Rayleigh fading or orthogonal pilots. We considered cell-free massive multiple-input multiple-output (CF mMIMO) and linear precoding to achieve mURLLC under Nakagami-m fading. Firstly, we establish a channel state information (CSI) error model with pilot contamination and analyze statistical properties of estimates and errors for Nakagami-m fading channels. Secondly, in conjunction with maximum ratio transmission (MRT) precoding, we derive the signal-to-interference-plus-noise ratio (SINR) of the downlink of the CF mMIMO system under Nakagami-m fading, which is meaningful but never mentioned in the literature. Then, we calculate the error probability (EP) with the finite blocklength and reveal interaction relationships between different indicators. The simulation results indicate that under pilot contamination, imperfect CSI, and Nakagami-m fading, mURLLC can be achieved by optimizing system configurations.

Keywords: Massive ultra-reliable and low latency communications (mURLLC) · Nakagami-m fading · Finite blocklength (FBL) · Cell-free massive multiple-input multiple-output (CF mMIMO) · Pilot sharing · Downlink precoding

This work was supported by the National Natural Science Foundation of China under grant 62001264, the Fundamental Research Funds for the Central Universities, and the Beijing Institute of Technology Research Fund Program for Young Scholars.

F. Gao et al. (Eds.): Chinacom 2023, LNICST 590, pp. 315–330, 2024.
https://doi.org/10.1007/978-3-031-67162-3_20

1 Introduction

Ultra-reliable and low latency communications (URLLC) requires latency less than 0.5 ms and reliability higher than 99.999%, making it one of the most challenging services [1–3]. As the scale of Internet of Things (IoT) devices grows dramatically, enhancing the access performance of the system is unavoidable to accommodate large-scale access for various requirements. Combining URLLC with large-scale access, massive ultra-reliable and low latency communications (mURLLC) can meet the future communication needs of IoT [4], which is a "killer" application scenario that can fully reflect the 6G revolution [5,6]. Nevertheless, tackling the three requirements of reliability, latency, and large-scale access to achieve mURLLC is fundamentally quite intractable [1]. In addition, how to design the mURLLC system architecture and transmission model to establish a high-density network with sufficiently large coverage is also a hot topic.

In order to cope with the above challenges, it is generally considered to adopt a new type of cell-free architecture to centrally process the signals of a substantial amount of access points (APs). It is more suitable for various heterogeneous communication scenarios of future wireless networks than the traditional cellular architecture. At the same time, massive multiple-input multiple-output (mMIMO) has kept vigorously developing in recent years, and has shown significant advantages in peak rate enhancement, spectral efficiency and reliability. Cell-free mMIMO (CF mMIMO), which combines mMIMO with the cell-free architecture, has the potential to meet the additional diverse needs of mURLLC. In addition, as a special distributed mMIMO system, CF mMIMO greatly shortens the distance from users to APs in traditional cells with a strong ability to resist fading, and shows a lot of promise to be one of the primary 6G technologies [7].

At present, relevant scholars have carried out research on URLLC and mURLLC based on CF mMIMO. Based on finite blocklength (FBL) information theory and stochastic network calculus (SNC) analysis tools, Zhang et al. [4] examined the CF mMIMO system to fulfill the strict requirements of mURLLC and accurately predicted the latency and error probability (EP) limitations. Nasir et al. [8] further optimized the downlink system's data rate and energy efficiency, proving that the CF mMIMO system has great potential to support URLLC services. Zeng et al. [9] combined maximum-ratio combining detection with simple path loss decoding, and achieve energy-efficient massive URLLC over CF mMIMO. Peng et al. [10] calculated the lower bound (LB) of ergodic data rate of downlink under the condition of finite block length and lack of accurate channel state information (CSI).

1.1 Motivation

Although some scholars have combined CF mMIMO for analysis, most of them are still discussing URLLC [8,10,11]. How to overcome the mutual constraints of latency-reliability-connection density to realize mURLLC is still an intractable problem. At the same time, although mURLLC is mentioned in papers such as [4] and [9], they all make the assumption that there is not any pilot contamination.

In this case, the pilot signals of different users are orthogonal to each other, which is impractical when there are sufficient users but insufficient pilot resources.

Additionally, the majority of papers make the assumption that the channel undergoes Rayleigh fading, which is not applicable when there is a line of sight (LoS) component present. Nakagami-m distribution can closely resemble the Rician fading model and serves as a more broad fading model that includes Rayleigh fading as a specific example [12]. Many of the current results on the Wishart matrix may be used for analysis when the components of the channel matrix $\mathbf{H}$ have a Gaussian distribution [13]. Nevertheless, it is challenging to analyze the higher-order statistical characteristics of $\mathbf{H}$ and the statistical characteristics of the correlation matrix when the elements of $\mathbf{H}$ are not Gaussian distributed. The existing papers only consider the special case when the $\mathbf{H}$ matrix dimension is 2 under Nakagami-m fading [14]. It is worth further exploring how to analyze the performance of massive users.

This study aims to conduct channel estimation and downlink data transmission under Nakagami-m fading and to further examine how user volume, reliability, and latency perform in relation to each other under various system parameters.

1.2 Contributions

1. With imperfect CSI under the Nakagami-m fading, we establish a CSI error model with pilot contamination and analyze statistical properties of estimates and errors of channels.
2. Using maximum ratio transmission (MRT) precoding, we develop a closed-form formula of signal-to-interference-plus-noise ratio (SINR) of downlink of CF mMIMO system under the Nakagami-m fading.
3. Using FBL information theory, we analyze the connections between reliability, latency, a number of connected users, pilot length, and other factors. To enable mURLLC, we then improve the CF mMIMO system parameters.

The remainder of this paper is organized as follows. Section 2 describes the CF mMIMO system model. In Sect. 3, we analyze the estimation of the uplink channel under Nakagami-m fading. In Sect. 4, we derive the downlink SINR and EP with FBL information theory. Section 5 presents some numerical results and related analysis, and Sect. 6 concludes this paper.

Notations: $\mathbb{C}^{K\times N}$ collects $K\times N$ complex-valued matrices. $\mathbb{E}(\bullet)$ and $\mathbb{D}(\bullet)$ denote expectation and variance, respectively. $(\bullet)^T$ denote transpose. $(\bullet)^H$ denote transpose and conjugate. $\mathcal{CN}(0,1)$ denotes the zero-mean complex Gaussian distribution with variance 1.

2 System Model

We assume a time-division duplex (TDD) mode. A CF mMIMO system contains U single-antenna users and Q multi-antenna APs. It is noting that there are F

antennas on each AP as shown in Fig. 1. $g_{q,fu}$ is assumed to be the channel, which is between the user u and the fth antenna of the AP q, and it is specified as follows

$$g_{q,fu} = \sqrt{\beta_{q,u}} h_{q,fu} \tag{1}$$

where $h_{q,fu} = |h_{q,fu}| e^{j\varphi}$ and $\beta_{q,u}$ represent the small-scale and large-scale fading coefficients, respectively. $\beta_{q,u}$ is known at each AP. φ and $|h_{q,fu}|$ are phase and amplitude of $h_{q,fu}$. They follows uniformly distributed in $[0, 2\pi)$ and Nakagami-m distribution, respectively [12,14]. The probability density function of $|h_{q,fu}|$ is shown below

$$f_{|h_{q,fu}|}(x) = \frac{2}{\Gamma(m_{q,u})} \left(\frac{m_{q,u}}{\Omega_{q,u}}\right) m_{q,u} x^{2m_{q,u}-1} e^{-(m_{q,u}/\Omega_{q,u})x^2}, x \geq 0, m_{q,u} \geq \frac{1}{2} \tag{2}$$

where $\Omega_{q,u} = \mathbb{E}[|h_{q,fu}|^2]$ representing the average power, which is a parameter of the Nakagami-m distribution and m is an another parameter. The gamma function is represented by $\Gamma(\cdot)$. The Nakagami-m squared envelope $|h_{q,fu}|^2$ has a gamma distribution i.e. Gamma $(m_{q,u}, \Omega_{q,u}/m_{q,u})$. Note that $|h_{q,fu}|$ is independent of φ, so $\mathbb{E}[h_{q,fu}] = 0$. $\mathbf{g}_{q,u} = \sqrt{\beta_{q,u}} \mathbf{h}_{q,u} \triangleq [g_{q,1u}, g_{q,2u}, ..., g_{q,Fu}]^T$, $\mathbf{h}_{q,u} \triangleq [h_{q,1u}, h_{q,2u}, ..., h_{q,Fu}]^T$. We assume that the elements of $\mathbf{h}_{q,u}$ are uncorrelated and $\mathbf{h}_{q,u}$ and $\mathbf{h}_{q,u'}$ are independent, for $u \neq u'$.

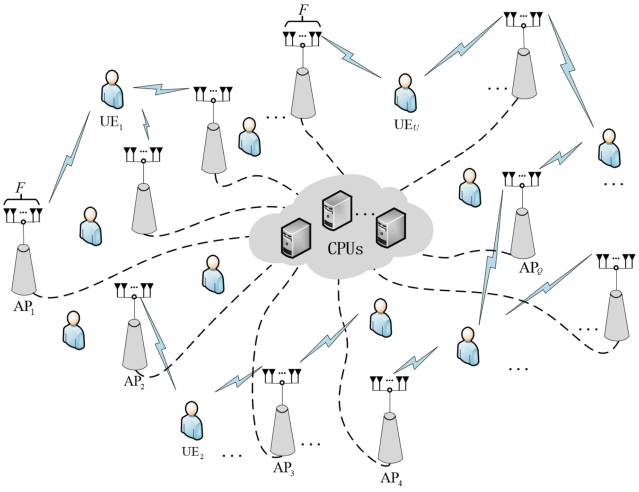


Fig. 1. A CF mMIMO system with multi-antenna distributed APs.

Here, pilot-assisted channel estimation (PACE) is taken to obtain CSI in the uplink. Then, APs use the channel estimates and apply MRT processing to transmit data to every users in the same time-frequency resource. $N = \tau_p + \tau_d$ is defined as the available number of channel uses, consisting of two parts τ_p and τ_d,

which are used for uplink channel estimation and downlink data transmission, respectively.

3 Uplink Channel Estimation

Initially, the user simultaneously sends a pilot signal to each AP in the system for uplink channel estimate. The length of the orthogonal pilot is denoted as τ_p. Considering the large number of users, we assume that $U \geq \tau_p$. $\mathbf{x}_{i_u}^p \in \mathbb{C}^{\tau_p \times 1}$ and $\mathbf{x}_{i_v}^p \in \mathbb{C}^{\tau_p \times 1}$ are the pilot sequence sent by user u and user v, which satisfy the relationship as follows.

$$\left(\mathbf{x}_{i_u}^p\right)^H \mathbf{x}_{i_v}^p = \begin{cases} 0, v \notin \mathcal{S}_u, \\ \tau_p, v \in \mathcal{S}_u, \end{cases} \tag{3}$$

where, i_u and $i_v \in \{1, 2, ..., \tau_p\}$ represent the index values of pilots for user u and user v, and $\mathcal{S}_u$ is the set of users who use the pilots used by user u. The pilot signal received at an AP q is

$$\begin{aligned} \mathbf{Y}_q^p &= \sqrt{\bar{p}}\mathbf{G}_q\mathbf{X}^p + \mathbf{Z}_q^p \\ &= \sqrt{\bar{p}}\left[\mathbf{g}_{q,1}, \mathbf{g}_{q,2}, ..., \mathbf{g}_{q,U}\right]\left[\left(\mathbf{x}_{i_1}^p\right), \left(\mathbf{x}_{i_2}^p\right), ..., \left(\mathbf{x}_{i_U}^p\right)\right]^T + \mathbf{Z}_q^p \\ &= \sqrt{\bar{p}}\sum_{j=1}^{U}\mathbf{g}_{q,j}\left(\mathbf{x}_{i_j}^p\right)^T + \mathbf{Z}_q^p \end{aligned} \tag{4}$$

where, $\bar{p}$ represents the average uplink transmission power, which is the same for each user. $\mathbf{G}_q = [\mathbf{g}_{q,1}, \mathbf{g}_{q,2}, ..., \mathbf{g}_{q,U}] \in \mathbb{C}^{F \times U}$ and $\mathbf{X}^p = \left[\left(\mathbf{x}_{i_1}^p\right), \left(\mathbf{x}_{i_2}^p\right), ..., \left(\mathbf{x}_{i_U}^p\right)\right]^T \in \mathbb{C}^{U \times \tau_p}$ are defined as a channel matrix between AP q and all users and a pilot signal matrix, respectively. $[\mathbf{G}_q]_{fu} = g_{q,fu} = \sqrt{\beta_{q,u}}h_{q,fu}$. $\mathbf{Z}_q^p \in \mathbb{C}^{F \times \tau_p}$ is a noise matrix in which the elements follow a complex Gaussian distribution with a mean of 0 and a variance of 1, and the elements are independent of each other. We can obtain the expression for the estimated channel by the minimum mean square error (MMSE) method, as given by [15,16]

$$\begin{aligned} \hat{\mathbf{g}}_{q,u} &= c_{q,u}\mathbf{Y}_q^p\left(\mathbf{x}_{i_u}^p\right)^* \\ &= c_{q,u}\left[\sqrt{\bar{p}}\sum_{j=1}^{U}\mathbf{g}_{q,j}\left(\mathbf{x}_{i_j}^p\right)^T + \mathbf{Z}_q^p\right]\left(\mathbf{x}_{i_u}^p\right)^* \\ &= c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_u}\mathbf{g}_{q,v}\tau_p + c_{q,u}\mathbf{Z}_l^p\left(\mathbf{x}_{i_u}^p\right)^* \\ &= c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_u}\mathbf{g}_{q,v}\tau_p + \bar{\mathbf{w}}_{q,u}, \end{aligned} \tag{5}$$

where, $c_{q,u} = \frac{\sqrt{\bar{p}}\beta_{q,u}}{\tau_p \sum\limits_{v\in\mathcal{S}_u}\bar{p}\beta_{q,v}+1}$, $\bar{\mathbf{w}}_{q,u} = c_{q,u}\mathbf{Z}_q^p\left(\mathbf{x}_{i_u}^p\right)^* \sim \mathcal{CN}\left(\mathbf{0}, c_{q,u}^2\tau_p\mathbf{I}_F\right)$. Considering that $g_{q,fu} = \sqrt{\beta_{q,u}}h_{q,fu}$ and $\mathbb{E}\left[h_{q,fu}\right] = 0$, we can deduce that $\mathbb{E}\left[\hat{\mathbf{g}}_{q,u}\right] = 0$

and $\mathbb{D}\left[\hat{\mathbf{g}}_{q,u}\right] = \lambda_{q,u}\mathbf{I}_F$, $\lambda_{q,u} = c_{q,u}^2\tau_p\left(\bar{p}\tau_p\sum\limits_{v\in\mathcal{S}_u}\beta_{q,v}\Omega_{q,v}+1\right)$. $\tilde{\mathbf{g}}_{q,u} = \mathbf{g}_{q,u} - \hat{\mathbf{g}}_{q,u}$ is the estimation error with $\mathbb{E}\left[\tilde{\mathbf{g}}_{q,u}\right] = 0$ and $\mathbb{D}\left[\tilde{\mathbf{g}}_{q,u}\right] = \left(\beta_{q,u}\Omega_{q,u} - \lambda_{q,u}\right)\mathbf{I}_F$.

Proof. See Appendix A.

Under the influence of pilot contamination, $\hat{\mathbf{g}}_{q,u} = \frac{c_{q,u}}{c_{q,v}}\hat{\mathbf{g}}_{q,v} = \frac{\beta_{q,u}}{\beta_{q,v}}\hat{\mathbf{g}}_{q,v}$ with $v \in \mathcal{S}_u$, $v \neq u$. This indicates that there is a correlation between $\hat{\mathbf{g}}_{q,u}$ and $\hat{\mathbf{g}}_{q,v}$ after using the same pilots.

4 Downlink Data Transmission and EP

4.1 Precoding and Received SINR

For downlink transmission, all APs send signals to U users simultaneously. Let $s(r) = [s_1(r), ..., s_U(r)]^T$ be the rth symbols intended for U users and $\mathbb{E}\left[s(r)s(r)^H\right] = \mathbf{I}_U$. After MRT processing, the precoded signal vector is

$$\mathbf{x}_q(r) = \sum_{u=1}^{U}\sqrt{p_{q,u}}\mathbf{w}_{q,u}s_u(r), \tag{6}$$

where $\mathbf{w}_{q,u} = \hat{\mathbf{g}}_{q,u}/\sqrt{\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^2\right]} = \hat{\mathbf{g}}_{q,u}/\sqrt{F\lambda_{q,u}}$ is the MRT precoding vector. $p_{q,u}$ is the downlink transmission power allocated by AP q to user u with $\sum_{u=1}^{U}p_{q,u} \leq \rho$, where ρ is maximum downlink transmit power. Then, we can obtain the downlink data signal at user u

$$\begin{aligned}
y_u(r) &= \sum_{q=1}^{Q}\mathbf{g}_{q,u}^H\mathbf{x}_q(r) + z_u(r) \\
&= \sum_{q=1}^{Q}\left[\sqrt{p_{q,u}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}s_u(r) + \sum_{j=1,j\neq u}^{U}\sqrt{p_{q,j}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,j}s_j(r)\right] + z_u(r), \\
&= \underbrace{\mathbb{E}[\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]}_{A_1u}\, s_u(r) \\
&+ \underbrace{\left\{\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,u} - \mathbb{E}[\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]\right\}}_{A_2u}s_u(r) \\
&+ \sum_{j=1,j\neq u}^{U}\underbrace{\sum_{q=1}^{Q}\sqrt{p_{q,j}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,j}}_{A_3uj}s_j(r) + z_u(r),
\end{aligned} \tag{7}$$

where A_1u, A_2u and A_3uj represent the coherent precoding gain, precoding gain uncertainty, and multi-user interference, respectively. $z_u(r)$ is Gaussian noise and $z_u(r) \sim \mathcal{CN}(0,1)$. In this case, the SINR of user u is given by

$$\gamma_u = \frac{|A_1 u|^2}{\mathbb{E}[|A_2 u|^2] + \sum\limits_{j=1, j\neq u}^{U} \mathbb{E}[|U A_3 uj|^2] + \mathbb{E}[|Z I_u|^2]}, \tag{8}$$

According to [15], γ_u can also be written as

$$\begin{aligned}\gamma_u &= \frac{\left|\sum\limits_{q=1}^{Q} \sqrt{p_{q,u}}\mathbb{E}[\mathbf{g}_{q,u}^H \mathbf{w}_{q,u}]\right|^2}{\sum\limits_{j=1}^{U} \mathbb{E}\left[\left|\sum\limits_{q=1}^{Q} \sqrt{p_{q,j}}\mathbf{g}_{q,u}^H \mathbf{w}_{q,j}\right|^2\right] - \left|\sum\limits_{q=1}^{Q} \sqrt{p_{q,u}}\mathbb{E}[\mathbf{g}_{q,u}^H \mathbf{w}_{q,u}]\right|^2 + 1} \\ &= \frac{F\left(\sum\limits_{q=1}^{Q} \sqrt{p_{q,u}\lambda_{q,u}}\right)^2}{A_u - F\left(\sum\limits_{q=1}^{Q} \sqrt{p_{q,u}\lambda_{q,u}}\right)^2 + 1}\end{aligned} \tag{9}$$

where $A_u = A_{u1} + A_{u2}$,

$$A_{u1} = \sum_{j\in\mathcal{S}_u} \left\{ \sum_{q=1}^{Q} p_{q,u} \frac{\beta_{q,j}^2}{\beta_{q,u}^2 \lambda_{q,u}} \left[\Delta_{q,u} + (F-1)\lambda_{q,u}^2\right] + \sum_{q=1}^{Q} \sum_{q'=1, q'\neq q}^{Q} \sqrt{p_{q,j} p_{q',j}} \frac{\beta_{q,j}\beta_{q',j}}{\beta_{q,u}\beta_{q',u}} \sqrt{F^2 \lambda_{q,u}\lambda_{q',u}} \right\}, \tag{10}$$

$$A_{u2} = \sum_{j=1}^{U} \sum_{q=1}^{Q} p_{q,j} \left(\beta_{q,u} - \lambda_{q,u}\right), \tag{11}$$

$$\begin{aligned}\Delta_{q,u} &= \mathbb{E}\left[\left|\hat{g}_{q,fu}\right|^4\right] \\ &= c_{q,u}^4 \bar{p}^2 \tau_p^4 \left(\sum_{v\in\mathcal{S}_u} \left(\frac{\Omega_{q,v}}{m_{q,v}}\right)^2 \frac{\Gamma(m_{q,v}+2)}{\Gamma(m_{q,v})}\right) \\ &+ c_{q,u}^4 \bar{p}^2 \tau_p^4 \left(2 \sum_{v\in\mathcal{S}_u} \sum_{\substack{v'\in\mathcal{S}_u, \\ v'\neq v}} \beta_{q,v}\beta_{q',v}\Omega_{q,v}\Omega_{q',v} + 4 c_{q,u}^4 \bar{p} \tau_p^3 \sum_{v\in\mathcal{S}_u} \beta_{q,v}\Omega_{q,u}\right).\end{aligned} \tag{12}$$

Proof. See Appendix B.

4.2 EP Under Finite Blocklength

We concentrate on FBL, which is suitable for URLLC. Under the aforementioned theoretical tools, the achievable data rate can be represented as [10,17]

$$R_u \approx (1-\eta)\, C(\gamma_u) - \sqrt{\frac{(1-\eta)\, V(\gamma_u)}{N}}\bar{Q}^{-1}(\varepsilon_u) \tag{13}$$

where $\eta = \frac{\tau_p}{N}$, $N = Bt_u$, t_u is the transmission latency of uth user, γ_u is the SINR of uth user, and ε_u is the EP of uth user. $C(\gamma) = \frac{\ln(1+\gamma)}{\ln 2}$, $V(\gamma) = \frac{\gamma(2+\gamma)}{[(1+\gamma)\ln 2]^2}$, and $\bar{Q}(x) = \frac{1}{\sqrt{2\pi}}\int_x^\infty \mathrm{e}^{(-\frac{y^2}{2})}dy$. Because $R_u = \varpi/(N-\tau_p)$, where ϖ represents the packet size, the EP obtained based on FBL theory is written as

$$\varepsilon_u = \bar{Q}\left(\sqrt{\frac{N}{(1-\eta)\, V(\gamma_u)}}\left((1-\eta)\, C(\gamma_u) - \frac{\varpi}{N-\tau_p}\right)\right). \tag{14}$$

5 Numerical Results and Analysis

We present theoretical and simulation data in this part to support the above-mentioned findings.

In a 2 × 2 km square, users and APs are dispersed uniformly and independently in the simulation scenario under consideration. Assumed to be 5*percent* is the activation probability of potential users. For large-scale fading, $\beta_{q,u} = \mu_p(d_{q,u}/d_0)^{-\alpha_p}$. The specific definitions and values of μ_p, α_p, d_0 are shown in Table 1. Here, each AP distributes its downlink signal power equally for users and $p_{q,u} = \rho/U$.

Table 1. Simulation parameter

Parameter	Value
Number of potential users, U	1000–2000
Uplink average TP (dBm)	20
Downlink power of each AP (W)	1
Bandwidth, B (MHz)	[100, 200]
Packet size, ϖ (bit)	30
Path loss exponent, α_p	3
Minimum distance, d_0 (m)	20
Constant path loss, μ_p (dB)	−10
Noise power spectral density (dBm/Hz)	−170

With a reasonable adjustment of the Nakagami-m fading parameters, we were able to replicate three of the classic fading models, Rayleigh and Rician,

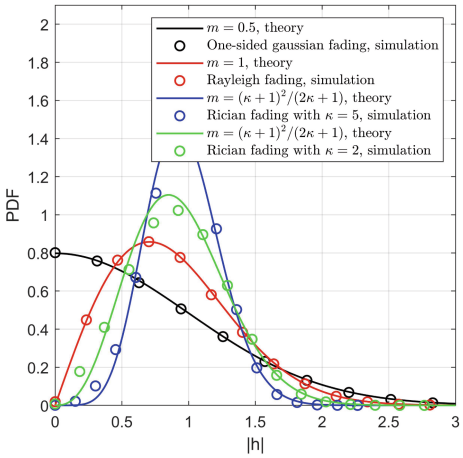


Fig. 2. PDF of the Nakagami-m RV.

and One-sided Gaussian distribution in Fig. 2. For example, when $m = 0.5$, the black circles representing the PDF of Nakagami-m distribution almost perfectly overlap on the black curve, which represents the one-side gaussian random variable generated by simulation. Similarly, by observing the blue and green curves and circles, we can find that the Nakagami-m distribution can fit the Rician distribution with different κ as the shape parameter κ of the Rician distribution changes when $m = (\kappa+1)^2/(2\kappa+1)$. The above results show the fitting ability of Nakagami-m distribution. Thus, our analysis under the Nakagami-m fading model has higher utility compared to that under the Rayleigh and Rician fading.

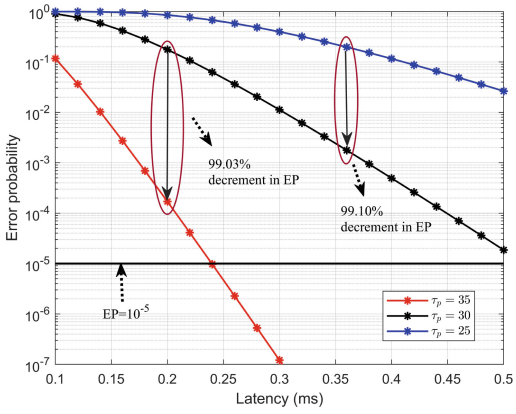


Fig. 3. EP vs. latency

We analyzes the variation law of EP with latency through simulation under several different pilot numbers, as shown in Fig. 3. Obviously, regardless of the value of τ_p, EP decreases gradually as the target latency increases. This suggests that there is a mutual constraint relationship between latency and reliability, and the performance of one can be appropriately reduced in exchange for the enhancement of the other. Moreover, comparing the three lines with different colors in the figure, it can be found that reliability can be enhanced by adding more orthogonal pilots. When $\tau_p = 25$, latency and reliability do not satisfy URLLC requirements simultaneously. However, when $\tau_p = 35$ and the given latency is greater than 0.24 ms, observing the red line, we can find that the reliability is always above 99.999%, which is due to the reduction of pilot contamination. Therefore, when the system performance requirements are strict, reliability and latency can be guaranteed simultaneously by increasing the orthogonal pilot resources.

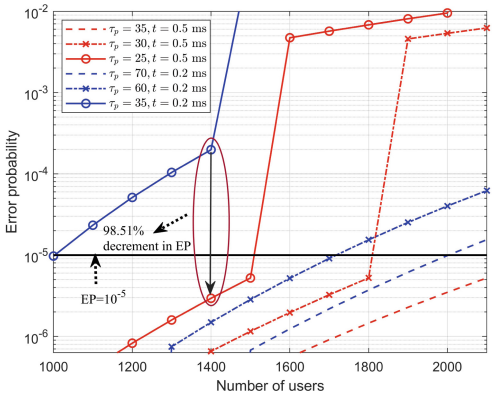


Fig. 4. EP vs. the number of users.

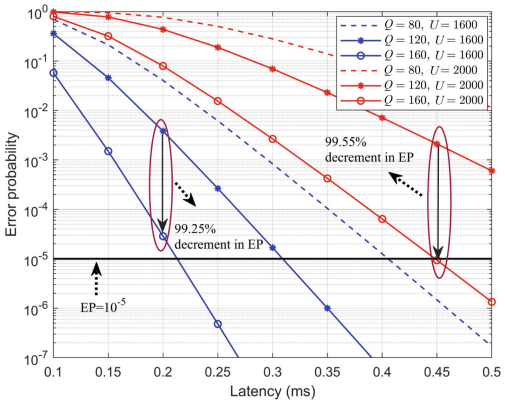


Fig. 5. EP vs. latency for different numbers of APs and users.

Figure 4 illustrates the interaction between reliability and the number of users. By observing the trend of the lines in the figure, it can be concluded that when the system accommodates more access users, the EP increases. This means that there is a mutually restrictive relationship between reliability and the number of connected users. We note that at $\tau_p = 25$, $t = 0.5$ ms, when the total number of households increases from 1500 to 1600, there is a jump segment with a large amplitude (from meeting the reliability requirement to not meeting the requirement) in EP. Similarly, at $\tau_p = 25$, $t = 0.5$ ms, EP also experienced a jump as the number of potential users increased from 1800 to 1900. These indicate that the system can only support a certain number of users at a particular latency and reliability. At this point, the three indicators of reliability, latency, and number of users can be guaranteed to be optimal simultaneously to achieve mURLLC. In addition, by increasing the number of pilots, reliability can be ensured under more stringent latency requirements.

By varying the quantity of APs, we investigate the influences between reliability and latency under various APs deployment situations in Fig. 5 in order to enhance system transmission performance to the greatest extent feasible with constrained antennas. Both the red and blue lines have the highest dashed line and the lowest solid line with a circle. This phenomenon means that under certain latency conditions, the more sufficient the AP, the higher the reliability of the system. For example, when $t = 0.45$ ms, the EP drops by 99.55% when Q rises from 120 to 160. In a similar vein, the two blue lines at the bottom of the graph show that, at $U = 1600$, $t = 0.2$ ms, an increase in APs can greatly enhance system reliability, resulting in a 99.25% improvement in reliability performance. This is because a denser deployment of APs results in a shorter path between users and them, which influences the large-scale fading coefficient.

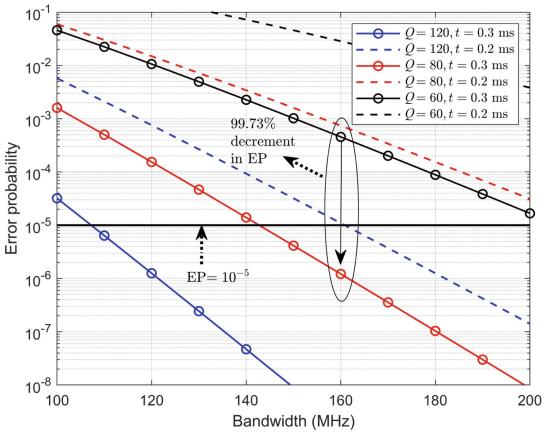
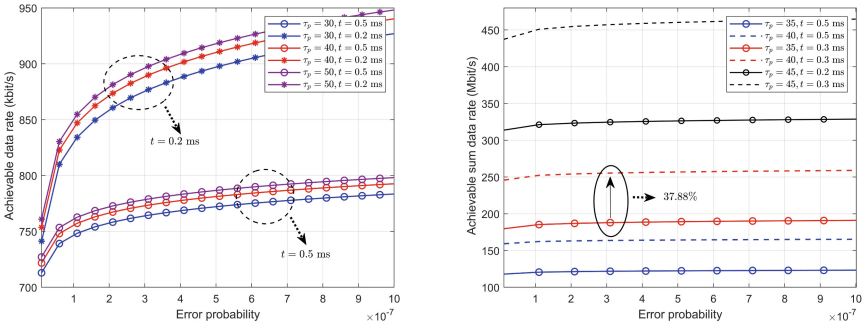


Fig. 6. EP vs. bandwidth

In addition, regardless of whether $U = 1600$ or $U = 2000$, users are able to meet the performance requirements under certain conditions. This shows that mURLLC can be implemented by adjusting the system parameter configuration (such as the AP numbers) to achieve the performance requirements under different numbers of users.

Furthermore, the Fig. 6 illustrates the relationship between EP and bandwidth in various AP deployment circumstances. On the one hand, every line on the graph indicates a declining trend as bandwidth increases, indicating that bandwidth resources can be utilized in exchange for reliability. For instance, if the deployment scenario of APs and the latency are fixed, then a suitable bandwidth increase can satisfy the lower EP. On the other hand, when the bandwidth is limited, we can compare the vertical axis of the curve and obtain that reliability can be guaranteed by adding more APs or reducing the latency requirement.



(a) User achievable data rate (b) System achievable data rate

Fig. 7. Achievable data rate vs. EP.

Figure 7 shows how the goal EP affects the achievable data rate, which reflects the trade-off of system between rate and reliability. Figure 7(a) is the single user considered and (b) is the whole system. It can be seen that when EP is given, the line containing the solid circle is above the line containing the hollow circle, indicating that the higher the target latency, the stricter the requirements for the achievable data rate. Meanwhile, when the latency and reliability are fixed, providing an appropriate number of orthogonal pilots can achieve greater data rates. For example, at EP equal to 10^{-7} and $t = 0.3$ ms, the sum data rate can be elevated by 35.88% when τ_p is increased from 35 to 40. These show that we can achieve different target performances by adjusting the system parameter configurations.

6 Conclusion

The impact of non-orthogonal pilots on the EP of downlink CF mMIMO systems during Nakagami-m fading was examined in this research. We derived the

statistical properties of the channel estimates and the downlink SINR under the linear precoding method. The EP was analytically evaluated by FBL information theory. Numerical analysis results showed that Nakagami-m fading can be fitted to Rayleigh, Rician, and one-sided Gaussian fading. In addition, notably, mURLLC is able to be attained by suitably modifying the quantity of APs, pilots, bandwidth, and numerous other system configurations.

A Proof of $\mathbb{E}\left[\hat{\mathbf{g}}_{l,k}\right]$, $\mathbb{D}\left[\hat{\mathbf{g}}_{l,k}\right]$, $\mathbb{E}\left[\tilde{\mathbf{g}}_{l,k}\right]$ and $\mathbb{D}\left[\tilde{\mathbf{g}}_{l,k}\right]$

We have,

$$\begin{aligned}\mathbb{E}\left[\hat{\mathbf{g}}_{q,u}\right] &= \mathbb{E}\left[c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_u}\mathbf{g}_{q,v}\tau_p + \bar{\mathbf{w}}_{q,u}\right] \\ &= c_{q,u}\sqrt{\bar{p}}\tau_p\sum_{v\in\mathcal{S}_u}\mathbb{E}[\mathbf{g}_{q,v}] + \mathbb{E}\left[\bar{\mathbf{w}}_{q,u}\right] = \mathbf{0}, \end{aligned} \tag{15}$$

$$\begin{aligned}\mathbb{D}\left[\hat{\mathbf{g}}_{q,u}\right] &= \mathbb{E}\left[\hat{\mathbf{g}}_{q,u}\hat{\mathbf{g}}_{q,u}^H\right] \\ &= \mathbb{E}\left[\left(c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_u}\mathbf{g}_{q,v}\tau_p + \bar{\mathbf{w}}_{q,u}\right)\left(c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_v}\mathbf{g}_{q,v}\tau_p + \bar{\mathbf{w}}_{q,u}\right)^H\right] \\ &= \mathbb{E}\left[c_{q,u}^2\bar{p}\tau_p^2\sum_{v\in\mathcal{S}_u}\mathbf{g}_{q,v}\mathbf{g}_{q,v}^H + c_{q,u}^2\bar{p}\tau_p^2\sum_{v\in\mathcal{S}_u}\sum_{v\in\mathcal{S}_u,v'\neq v}\mathbf{g}_{q,v}\mathbf{g}_{q,v'}^H + \bar{\mathbf{w}}_{q,u}\bar{\mathbf{w}}_{q,u}^H\right] \\ &\overset{(a)}{=} c_{q,u}^2\bar{p}\tau_p^2\sum_{v\in\mathcal{S}_u}\beta_{q,v}\Omega_{q,v}\mathbf{I}_F + 0\mathbf{I}_F + c_{q,u}^2\tau_p\mathbf{I}_F \\ &= c_{q,u}^2\tau_p\left(\bar{p}\tau_p\sum_{v\in\mathcal{S}_u}\beta_{q,v}\Omega_{q,v} + 1\right)\mathbf{I}_F \\ &\triangleq \lambda_{q,u}\mathbf{I}_F, \end{aligned} \tag{16}$$

where, the condition for equation (a) to hold is that the channels between different users and the same AP are independent of each other. The channel estimation error can be expressed as $\tilde{g}_{q,u} = g_{q,u} - \hat{g}_{q,u}$ and is independent of the estimate $\tilde{g}_{q,u}$. Therefore, $\mathbb{E}\left[\tilde{\mathbf{g}}_{q,u}\right] = \mathbb{E}\left[\mathbf{g}_{q,u} - \hat{\mathbf{g}}_{q,u}\right] = 0$, $\mathbb{D}\left[\tilde{\mathbf{g}}_{q,u} + \hat{\mathbf{g}}_{q,u}\right] = \mathbb{D}\left[\tilde{\mathbf{g}}_{q,u}\right] + \mathbb{D}\left[\hat{\mathbf{g}}_{q,u}\right] = \mathbb{D}\left[\mathbf{g}_{q,u}\right]$, $\mathbb{D}\left[\tilde{\mathbf{g}}_{q,u}\right] = \mathbb{D}\left[\mathbf{g}_{q,u}\right] - \mathbb{D}\left[\hat{\mathbf{g}}_{q,u}\right] = \left(\beta_{q,u}\Omega_{q,u} - \lambda_{q,u}\right)\mathbf{I}_F$. The proof of the expectation and variance of channel estimates and estimation errors is completed.

B Proof of Downlink SINR

In this part, we provide proof of (9). According to [15], the further simplified expression of SINR can be expressed as

$$\gamma_u = \frac{\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}[\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]\right|^2}{\sum_{j=1}^{U}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,j}\right|^2\right]-\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}[\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]\right|^2+1} \tag{17}$$

where,

$$\begin{aligned}&\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbb{E}[\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]\right|^2\\ =&\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbb{E}\left[\left(\hat{\mathbf{g}}_{q,u}^H+\tilde{\mathbf{g}}_{q,u}^H\right)\mathbf{w}_{q,u}\right]\right|^2=\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbb{E}\left[\hat{\mathbf{g}}_{q,u}^H\mathbf{w}_{q,u}\right]\right|^2\end{aligned} \tag{18}$$

$\mathbf{w}_{q,u}=\frac{\hat{\mathbf{g}}_{q,u}}{\sqrt{\mathbb{E}[||\hat{\mathbf{g}}_{q,u}||^2]}}=\frac{\hat{\mathbf{g}}_{q,u}}{\sqrt{F\lambda_{q,u}}}$ when using MRT precoding. Furthermore, we have

$$\begin{aligned}&\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbb{E}[\mathbf{g}_{q,u}^H\mathbf{w}_{q,u}]\right|^2=\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\mathbb{E}\left[\hat{\mathbf{g}}_{q,u}^H\frac{\hat{\mathbf{g}}_{q,u}}{\sqrt{F\lambda_{q,u}}}\right]\right|^2\\ =&\left|\sum_{q=1}^{Q}\sqrt{p_{q,u}}\frac{1}{\sqrt{F\lambda_{q,u}}}\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^2\right]\right|^2=F\left(\sum_{q=1}^{Q}\sqrt{p_{q,u}\lambda_{q,u}}\right)^2\end{aligned} \tag{19}$$

The term $\sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,j}\right|^2\right]$ is given by

$$\begin{aligned}\sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\mathbf{g}_{q,u}^H\mathbf{w}_{q,j}\right|^2\right]&=\sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\left(\hat{\mathbf{g}}_{q,u}^H+\tilde{\mathbf{g}}_{q,u}^H\right)\mathbf{w}_{q,j}\right|^2\right]\\ &\overset{(b)}{=}\sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\hat{\mathbf{g}}_{q,u}^H\mathbf{w}_{q,j}\right|^2\right]+\sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\tilde{\mathbf{g}}_{q,u}^H\mathbf{w}_{q,j}\right|^2\right]\\ &\triangleq A_{u1}+A_{u2}\end{aligned} \tag{20}$$

where the condition for equation (b) is that $\tilde{\mathbf{g}}_{q,u}$ is independent of $\hat{\mathbf{g}}_{q,u}$. Substituting $\mathbf{w}_{q,j} = \frac{\hat{g}_{q,j}}{\sqrt{F\lambda_{q,j}}}$ into (18), we obtain

$$\begin{aligned}
A_{u1} &= \sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\hat{\mathbf{g}}_{q,u}^{H}\mathbf{w}_{q,j}\right|^2\right] = \sum_{j\in\mathcal{S}_u}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\hat{\mathbf{g}}_{q,u}^{H}\frac{\hat{\mathbf{g}}_{q,j}}{\sqrt{F\lambda_{q,j}}}\right|^2\right] \\
&= \sum_{j\in\mathcal{S}_u}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\frac{1}{\sqrt{F\lambda_{q,j}}}\frac{\beta_{q,j}}{\beta_{q,u}}||\hat{\mathbf{g}}_{q,u}||^2\right|^2\right] \\
&= \sum_{j\in\mathcal{S}_u}^{U}\sum_{q=1}^{Q}p_{q,j}\frac{\beta_{q,j}^2}{\beta_{q,u}^2}\frac{1}{F\lambda_{q,j}}\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^4\right] \\
&+ \sum_{j\in\mathcal{S}_u}^{U}\sum_{q'=1}^{Q}\sum_{q'=1,q'\neq l}^{Q}\sqrt{p_{q,j}p_{q',j}}\frac{\beta_{q,j}\beta_{q',j}}{\beta_{q,u},\beta_{q',u}}\frac{1}{\sqrt{F^2\lambda_{q,j}\lambda_{q',j}}}\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^2\right]\mathbb{E}\left[\mathbb{E}||\hat{\mathbf{g}}_{q',u}||^2\right],
\end{aligned} \tag{21}$$

where,

$$\begin{aligned}
\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^4\right] &= \mathbb{E}\left[\left(||\hat{\mathbf{g}}_{q,u}||^2\right)^2\right] = \mathbb{E}\left[\left(\sum_{f=1}^{F}\left|\hat{g}_{q,fu}\right|^2\right)^2\right] \\
&= \sum_{f=1}^{F}\mathbb{E}\left[\left|\hat{g}_{q,fu}\right|^4\right] + \sum_{m=1}^{F}\sum_{f'=1,f\neq 1}^{F}\mathbb{E}\left[\left|\hat{g}_{q,fu}\right|^2\right]\mathbb{E}\left[\left|\hat{g}_{q,f'u}\right|^2\right] \\
&= \sum_{f=1}^{F}\mathbb{E}\left[\left|c_{q,u}\sqrt{\bar{p}}\sum_{v\in\mathcal{S}_u}g_{q,v}\tau_p + w_{q,u}\right|^4\right] + F\left(F-1\right)\lambda_{q,u}^2 \\
&= F\Delta_{q,u} + F\left(F-1\right)\lambda_{q,u}^2,
\end{aligned} \tag{22}$$

The specific expression of $\Delta_{q,u}$ is shown in (12). Due to space constraints, we have omitted some cumbersome steps, but it is worth noting that $|h_{q,fu}|^4$ follows the generalized gamma distribution with parameters $p1 = 1/2$, $d1 = m_{q,u}/2$ and $\mathbb{E}\left[|h_{q,fu}|^4\right] = a1\frac{\Gamma((d1+1)/p1)}{\Gamma(d1/p1)} = \left(\frac{\Omega_{q,u}}{m_{q,u}}\right)^2\frac{\Gamma(m_{q,u}+2)}{\Gamma(m_{q,u})}$ when $|h_{q,u}|^2 \sim$ Gamma $(m_{q,u}, \Omega_{q,u}/m_{q,u})$. Substituting (20) and $\mathbb{E}\left[||\hat{\mathbf{g}}_{q,u}||^2\right]\left[||\hat{\mathbf{g}}_{q',u}||^2\right] = F^2\lambda_{q,u}\lambda_{q',u}$ into (19), we obtain A_{u1} in (10). Furthermore, $A_{u2} = \sum_{j=1}^{U}\mathbb{E}\left[\left|\sum_{q=1}^{Q}\sqrt{p_{q,j}}\tilde{\mathbf{g}}_{q,u}^{H}\mathbf{w}_{q,j}\right|^2\right] = \sum_{j=1}^{U}\mathbb{E}\left[\sum_{q=1}^{Q}p_{q,j}\tilde{\mathbf{g}}_{q,u}^{H}\mathbf{w}_{q,j}\mathbf{w}_{q,j}^{H}\tilde{\mathbf{g}}_{q,u}\right] + 0 = \sum_{j=1}^{U}\sum_{q=1}^{Q}p_{q,j}\left(\beta_{q,u}\Omega_{q,u} - \lambda_{q,u}\right)$. Then, we obtain γ_u in (9).

References

1. Liu, Y., Deng, Y., Elkashlan, M., Nallanathan, A., Karagiannidis, G.K.: Optimization of grant-free NOMA with multiple configured-grants for mURLLC. IEEE J. Sel. Areas Commun. **40**(4), 1222–1236 (2022)
2. Popovski, P., Stefanovic, C., Nielsen, J.J., Carvalho, E.D., Bana, A.S.: Wireless access in ultra-reliable low-latency communication (URLLC). IEEE Trans. Commun. **67**(8), 5783–5801 (2019)
3. Wu, F., Chen, L., Zhao, N., Chen, Y., Yu, F.R., Wei, G.: Computation over multi-access channels: multi-hop implementation and resource allocation. IEEE Trans. Commun. **69**(2), 1038–1052 (2021)
4. Zhang, X., Wang, J., Poor, H.V.: Statistical delay and error-rate bounded QoS provisioning for mURLLC over 6G CF M-MIMO mobile networks in the finite blocklength regime. IEEE J. Sel. Areas Commun. **39**(3), 652–667 (2021)
5. You, X., et al.: Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. SCIENCE CHINA Inf. Sci. **64**(1), 74 (2021)
6. Saad, W., Bennis, M., Chen, M.: A vision of 6G wireless systems: applications, trends, technologies, and open research problems. IEEE Network **34**(3), 134–142 (2019)
7. Elhoushy, S., Ibrahim, M., Hamouda, W.: Cell-free massive MIMO: a survey. IEEE Commun. Surv. Tutor. **24**(1), 492–523 (2021)
8. Nasir, A.A., Tuan, H.D., Ngo, H.Q., Duong, T.Q., Poor, H.V.: Cell-free massive MIMO in the short blocklength regime for URLLC. IEEE Trans. Wireless Commun. **20**(9), 5861–5871 (2021)
9. Zeng, J., Wu, T., Song, Y., Zhong, Y., Lv, T., Zhou, S.: Achieving energy-efficient massive URLLC over cell-free massive MIMO. IEEE Internet Things J. (2023)
10. Peng, Q., Ren, H., Pan, C., Liu, N., Elkashlan, M.: Resource allocation for cell-free massive MIMO-enabled URLLC downlink systems. IEEE Trans. Veh. Technol. **72**(6), 7669–7684 (2023)
11. Peng, Q., Ren, H., Pan, C., Liu, N., Elkashlan, M.: Resource allocation for uplink cell-free massive MIMO enabled URLLC in a smart factory. IEEE Trans. Commun. **71**(1), 553–568 (2022)
12. Zhong, C., Matthaiou, M., Huang, A., Zhang, Z.: On the sum rate of MIMO Nakagami-m fading channels with linear receivers. IEEE Trans. Wireless Commun. **11**(10), 3651–3659 (2012)
13. Tulino, A.M., Verdú, S.: Random matrix theory and wireless communications. Found. Trends Commun. Inf. Theory **1**(1), 1–163 (2004)
14. Fraidenraich, G., Lévêque, O., Cioffi, J.M.: On the MIMO channel capacity for the Nakagami-m channel. IEEE Trans. Inf. Theory **54**(8), 3752–3757 (2008)
15. Interdonato, G., Karlsson, M., Björnson, E., Larsson, E.G.: Local partial zero-forcing precoding for cell-free massive MIMO. IEEE Trans. Wireless Commun. **19**(7), 4758–4774 (2020)
16. Interdonato, G., Karlsson, M., Björnson, E., Larsson, E.G.: Downlink spectral efficiency of cell-free massive MIMO with full-pilot zero-forcing. In: 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Anaheim, pp. 1003–1007. IEEE (2018)
17. Polyanskiy, Y., Poor, H.V., Verdú, S.: Channel coding rate in the finite blocklength regime. IEEE Trans. Inf. Theory **56**(5), 2307–2359 (2010)