Statistical Delay and Error-Rate Bounded QoS Provisioning for mURLLC Over 6G CF M-MIMO Mobile Networks in the Finite Blocklength Regime

Xi Zhang, Fellow, IEEE, Jingqing Wang, Fellow, IEEE, and H. Vincent Poor, Life Fellow, IEEE

Abstract—In supporting the new 6G standard traffic services—massive ultra-reliable low-latency communications (mURLLC), several advanced techniques, including statistical delay-bounded quality-of-service (QoS) provisioning theory and finite blocklength coding (FBC), have been developed to upper-bound both delay and error-rate for time-sensitive multimedia applications. On the other hand, cell-free (CF) massive multi-input multi-output (m-MIMO), where a large number of distributed access points (APs) jointly serve a massive number of mobile devices using the same time-frequency resources, has emerged as one of the 6G key promising techniques to significantly improve various QoS performances for supporting mURLLC. However, it is challenging to statistically guarantee stringent mURLLC QoS requirements for transmitting multimedia traffic over CF m-MIMO and FBC based 6G wireless networks. To overcome these problems, we develop analytical models to precisely characterize the delay and error-rate bounded QoS performances while considering non-vanishing decode-error probability for CF m-MIMO based schemes. In particular, we develop FBC based system models and apply the Mellin transform to characterize arrival/service processes for our proposed CF m-MIMO modeling schemes. Then, we formulate and solve the delay violation probability minimization problem and obtain the closed-form solution of the optimal rate adaptation policy for each mobile user over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime. Our simulation results validate and evaluate our proposed schemes for statistical delay and error-rate bounded QoS provisioning.

Index Terms—Statistical delay and error-rate bounded QoS, 6G, mURLLC, CF m-MIMO, FBC, Mellin transform, SNC.

I. INTRODUCTION

Due to the stochastic nature of wireless fading channels, it is challenging to guarantee both reliability and low-latency requirements for delay-sensitive wireless multimedia services over 6G mobile wireless networks. Traditionally, researchers have developed a deterministic network calculus to derive explicit guarantees on the maximum delay for wireless data transmissions. However, deterministic network calculus is not sufficient for characterizing the wireless traffic due to time-varying and stochastic natures of wireless fading channels. Towards this end, the delay-bounded QoS theory [1] [4] and stochastic network calculus (SNC) [3] have been developed to analyze queueing behaviors based on the theories of large deviations and effective capacity. Particularly, the concept of statistical QoS provisioning [1] [4], in terms of effective capacity and delay-bounded violation probabilities, has been proposed to support time-sensitive wireless communications over 6G mobile wireless networks.

In addition, as a new and dominating 6G mobile-networks’ service class for time-sensitive traffic, massive ultra-reliable and low latency communications (mURLLC), which integrates URLLC with massive access, also known as massive machine type communications (mMTC), requires massive short-packet data communications to support time-sensitive 6G wireless multimedia services with high resource efficiency and low access latency [5]. This implies that the traditional Shannon’s theorem with infinite blocklength is no longer applicable in this regime. Towards this end, finite blocklength coding (FBC) [6] has been proposed to guarantee both latency and reliability requirements using short-packet data communications for supporting 6G wireless multimedia services. The maximum achievable coding rate using FBC over additive white Gaussian noise (AWGN) channels has been derived in [7]. The authors of [8] have derived the goodput over AWGN channels and the energy-efficiency spectral-efficiency tradeoff by using recent results on the non-asymptotic coding rate. The maximum achievable data rates using FBC over static MIMO based wireless fading channels with and without the knowledge of channel state information (CSI) have been derived in [9]. The authors of [10] have investigated different properties of channel codes for a given memoryless wireless channel with a non-vanishing error probability.

On the other hand, various advanced promising 6G techniques, such as the cell-free (CF) massive multi-input multi-output (m-MIMO) [11], have been designed to play a critically important role in supporting mURLLC as well as mMTC in terms of connecting massive numbers of mobile devices without imposing congestion. In particular, CF m-MIMO [12], where the geographically distributed APs jointly
serve a massive number of mobile devices using the same time-frequency resources, has been developed as a promising 6G network architecture for improving the reliability of massive access while reducing the co-channel interference caused by standard m-MIMO systems. Traditionally, optimal power control is performed at the central processing unit (CPU). However, centralized power-control strategies may jeopardize the system scalability and violate mURLLC requirements as the numbers of APs and mobile users grow significantly. The authors of [13] have proposed scalable and distributed power control policies for CF m-MIMO systems to achieve system scalability and mURLLC as the number of mobile users goes to infinity. The authors of [14] have developed a new framework for scalable CF m-MIMO systems, where the complexity and signalling at each AP is finite when connecting a massive number of mobile devices. The system scalability aspects of CF m-MIMO systems are analyzed in [15] and a solution is proposed for data processing, network topology, and power control. However, it is challenging to characterize the stochastic networking/queueing behaviors when being integrated with CF m-MIMO schemes while guaranteeing mURLLC under statistical delay and error-rate bounded QoS constraints in the finite blocklength regime. As a result, how to accurately upper-bound the delay violation probability while guaranteeing statistical delay and error-rate bounded QoS provisionings for supporting mURLLC still remains as a challenging and open problem for CF m-MIMO mobile wireless networks in the finite blocklength regime.

To effectively overcome the above-mentioned challenges, in this paper we propose to apply the Mellin transform to analytically model and characterize stochastic QoS performances in terms of both delay and error-rate for CF m-MIMO modeling schemes in the finite blocklength regime. In particular, we develop CF m-MIMO based system models across Rician wireless fading channels in the finite blocklength regime. Furthermore, we propose and analyze the delay-violation probability function by applying the Mellin transform over arrival/service processes, while taking into account the statistical delay and error-rate bounded QoS constraints. We also formulate and solve the delay-violation probability minimization problem for our proposed CF m-MIMO modeling schemes in the finite blocklength regime. Also conducted is a set of simulations to validate and evaluate our proposed schemes for statistical delay and error-rate bounded QoS provisioning over 6G CF m-MIMO mobile wireless networks.

The rest of this paper is organized as follows: Section II establishes FBC based CF m-MIMO based system models across Rician wireless channels. Section III derives and analyzes the Mellin transform over arrival/service processes as well as an upper bound on the delay violation probability in the finite blocklength regime. Section IV analyzes the delay performance and formulates and solves the delay violation probability minimization problem for statistical delay and error-rate bounded QoS provisioning in the finite blocklength regime. Section V evaluates and analyzes the system performance for our proposed FBC based CF m-MIMO schemes. The paper concludes with Section VI.

II. THE SYSTEM MODELS

Consider a CF m-MIMO network model, where each mobile user is served by coherent joint transmissions from all APs, as shown in Fig. 1. Assume that there are \( K_a \) randomly located APs over a large area and \( K_u \) mobile users. Assume that each AP is equipped with \( N_T \) antennas, while each mobile user is equipped with a single antenna. In addition, time division duplexing (TDD) mode is assumed to be operated over our proposed 6G CF m-MIMO mobile wireless networks. All APs are connected to a CPU through backhaul links. Define \( n_p \) as the number of channel uses for uplink pilot training symbols and \( n_d \) as the number of channel uses reserved for downlink data transmissions. Defining \( n \) as the total number of channel uses for both uplink pilot training and downlink data transmission phases, we have \( n = n_p + n_d \).

A. Massive MIMO Based Rician Wireless Fading Channel Model

Considering the propagation effects, the channel’s impulse response vector, denoted by \( g_{k,m} \in \mathbb{C}^{N_T \times 1} \), between mobile user \( m \) and AP \( k \) over Rician massive MIMO based wireless fading channel model can be characterized as follows:

\[
g_{k,m} = h_{k,m} \sqrt{\beta_{k,m}}
\]  

(1)

where \( \beta_{k,m} \) represents the large-scale propagation that includes pathloss and shadowing effects and \( h_{k,m} \) represents the small-scale multipath fading effect which can be modeled using Rician distribution [16] as follows:

\[
h_{k,m} = \sqrt{\frac{\kappa}{\kappa + 1}} \bar{h}_{k,m} + \sqrt{\frac{1}{\kappa + 1}} h_{k,m}
\]  

(2)

where \( h_{k,m} \) consists of the component \( \bar{h}_{k,m} \) representing the line of sight (LOS) signals and a Rayleigh distributed random component \( h_{k,m} \) representing the non-line-of-sight (NLOS) signals, and \( \kappa > 0 \) is the Rician factor. Note that \( \kappa = 0 \) corresponds to a Rayleigh fading channel, while \( \kappa \to \infty \) corresponds to non-fading channel. We can rewrite the
Rician massive MIMO based wireless fading channel model as follows:

\[ g_{k,m} = \sqrt{\frac{\kappa}{\kappa + 1}} \bar{g}_{k,m} + \sqrt{\frac{1}{\kappa + 1}} \hat{g}_{k,m}. \]  

(3)

where \( \bar{g}_{k,m} \triangleq T_{k,m} \beta_{k,m} \) and \( \hat{g}_{k,m} \triangleq h_{k,m}/\beta_{k,m} \).

**B. Uplink Pilot Training and Channel Estimation**

Define the pilot training sequence for mobile user \( m \) as \( \phi^m_\kappa = \left[ \phi^1_\kappa, \ldots, \phi^{N_p}_\kappa \right] \in \mathbb{C}^{1 \times N_p} \) and \( \| \phi^m_\kappa \| = 1 \), where \( \| \cdot \| \) denotes the Euclidean norm. During the uplink pilot training phase, we derive the received signal, denoted by \( \tilde{Y}^m_n \in \mathbb{C}^{N_t \times n_r} \), for transmitting \( N_p \) pilot data blocks from mobile user \( m \) to AP \( k \) as in the following equation:

\[ \tilde{Y}^m_n = G_k \phi^m_n + N_k. \]  

(4)

Then, by projecting the received signal \( \tilde{Y}^m_n \) onto \( \phi^m_n \), we obtain:

\[ \tilde{y}^m_n = \tilde{Y}^m_n \phi^m_n = \sqrt{N_p P_p} g_{k,m} \phi^m_n + \tilde{n}_k. \]  

(5)

where \( (\cdot)^H \) represents the conjugate transpose of a vector and \( \tilde{n}_k \triangleq N_k (\phi^m_n)^H \) is an independent and identically distributed (i.i.d.) Gaussian vector with zero mean and covariance \( I_{N_t} \). Denote by \( \tilde{G}_k \equiv \left[ g_{k,1}, \ldots, g_{k,K_u} \right] \) the NLOS component of the channel’s impulse response matrix from all AP \( k \) to all mobile users. Define \( \tilde{R}_{\tilde{G}_k} \equiv \mathbb{E} \left[ \tilde{G}_k \tilde{G}_k^H \right] = \text{diag}(\beta_{k,1}, \ldots, \beta_{k,K_u}) \) as the covariance matrix of \( \tilde{G}_k \), where \( \mathbb{E}[\cdot] \) is the expectation operation and \( \text{diag}(\cdot) \) represents the diagonal matrix. Then, considering the Rician wireless fading channels, we can derive the channel estimation for our proposed CF m-MIMO schemes as detailed in the following lemma.

**Lemma 1.** The minimum mean-squared error (MMSE) estimator, denoted by \( \hat{G}_k \), for the NLOS component of the Rician massive MIMO based wireless fading channel \( G_k \) from AP \( k \) to all mobile users is derived as follows:

\[ \hat{G}_k = \sqrt{\frac{N_p P_p}{\kappa + 1}} \tilde{R}_{\tilde{G}_k} \left( \frac{N_p P_p}{\kappa + 1} \tilde{R}_{\tilde{G}_k} + I_{N_t} \right)^{-1} \tilde{y}^n_k. \]  

(7)

**Proof:** Applying the MMSE estimator of \( \tilde{G}_k \) based on the observation of \( \tilde{y}^n_k \), we can obtain the following equation [17]:

\[ \hat{G}_k = \mathbb{E} \left[ \tilde{G}_k | \tilde{y}^n_k \right] = R_{\tilde{G}_k, \tilde{y}^n_k} \left( R_{\tilde{y}^n_k} \right)^{-1} \left( \tilde{y}^n_k - \mathbb{E} \left[ \tilde{y}^n_k \right] \right) + \mathbb{E} \left[ \tilde{G}_k \right] \]  

(8)

where \( R_{\tilde{G}_k, \tilde{y}^n_k} \) and \( R_{\tilde{y}^n_k} \) represent the covariance matrices, respectively, as follows:

\[ \left\{ \begin{array}{l} R_{\tilde{G}_k, \tilde{y}^n_k} = \mathbb{E} \left[ \tilde{G}_k (\tilde{y}^n_k)^H \right] = \frac{N_p P_p}{\kappa + 1} R_{\tilde{G}_k} \, ; \\ R_{\tilde{y}^n_k} = \mathbb{E} \left[ \tilde{y}^n_k (\tilde{y}^n_k)^H \right] = \frac{N_p P_p}{\kappa + 1} R_{\tilde{G}_k} + I_{N_t} \end{array} \right. \]  

(9)

Since \( \mathbb{E} \left[ \tilde{y}^n_k \right] \) and \( \mathbb{E} \left[ \tilde{G}_k \right] \) are equal to zero, we have

\[ \hat{G}_k = \sqrt{\frac{N_p P_p}{\kappa + 1}} \tilde{R}_{\tilde{G}_k} \left( \frac{N_p P_p}{\kappa + 1} \tilde{R}_{\tilde{G}_k} + I_{N_t} \right)^{-1} \tilde{y}^n_k \]  

(10)

which completes the proof of Lemma 1. 

**C. Downlink Finite-Blocklength Data Transmission**

Define \( P_m \) as the multiplexing order for mobile user \( m \), i.e., the number of information symbols simultaneously transmitted at the transmitter. As a result, we define a \( P_m \)-dimensional beamformer \( L_m \equiv I_{P_m} \otimes I_{N_k/P_m} \) for mobile user \( m \), where \( \otimes \) is the Kronecker product, \( I_{P_m} \) is the identity matrix of size \( P_m \), and \( I_{N_k/P_m} \) is the all one vector with length \( N_k/P_m \). We define the transmit signal matrix as \( X^n_m \equiv \left[ x^{(1)}_m, \ldots, x^{(n_m)}_m \right] \) and receive signal vector \( y^n_m \equiv \left[ y^{(1)}_m, \ldots, y^{(n_m)}_m \right] \). Based on the MMSE estimator matrix, denoted by \( \hat{G}_k \triangleq \left[ \hat{g}_{k,1}, \ldots, \hat{g}_{k,K_u} \right] \), obtained during the uplink pilot training phase, we can derive the transmitted signal, denoted by \( x^{(l)}_k \), for transmitting data block \( l \) from AP \( k \) to mobile user \( m \) by employing conjugate beamforming [11] as follows:

\[ x^{(l)}_k = W_k (\Sigma_k)^{1/2} s^{(l)}_m \, , \, l = 1, \ldots, n_d \]  

(12)

where \( s^{(l)}_m \) represents the \( l \)-th transmitted data block to mobile user \( m \), \( \Sigma_k \triangleq \text{diag}(\eta_{k,1}, \ldots, \eta_{k,K_u}) \) is the power allocation coefficient matrix, where \( \eta_{k,m} (m = 1, \ldots, K_u) \) is the power allocation coefficient for transmitting finite-blocklength data block \( l \) from AP \( k \) to mobile user \( m \), and \( W_k \) is the downlink precoder from AP \( k \) to all \( K_u \) mobile users, which is given by

\[ W_k \triangleq \hat{G}_k \left( \hat{G}_k^H \hat{G}_k \right)^{-1} L_m (\Xi_k)^{1/2} \]  

(13)

where \( \Xi_k = \text{diag}(\chi_1, \ldots, \chi_{K_u}) \) is the normalization matrix such that the columns of \( W_k \) have unit norm and the normalization variable \( \chi_k \) with \( k \in \{1, \ldots, K_u\} \) follows the central

\[ \chi_k \sim \text{CN}(0, 1). \]  

(14)
chi-square distribution with $(2\ell)$ degrees of freedom, where 
\[ \ell = K_a - K_u + 1. \] 
The probability density function (PDF) of \( \chi_k \) is given as follows [18]: 
\[ f_{\ell}(\chi_k) = \frac{1}{\Gamma(\ell)} \chi_k^{\ell-1} e^{-\chi_k} \] 
(14)

where \( \Gamma(\cdot) \) denotes the Gamma function. We have \( W_k = [w_{k,1}, \ldots, w_{k,K_u}] \), where \( w_{k,m} \) is the downlink precoder from AP \( k \) to mobile user \( m \) \( (m \in \{1, \ldots, K_u\}) \). In addition, the power control coefficients need to satisfy the following power constraint at each AP:
\[ \frac{1}{n_d} \sum_{l=1}^{n_d} \mathbb{E}\left[ \left\| x^{(l)}_k \right\|^2 \right] \leq P_d \] 
(15)

where \( P_d \) represents the average transmit power at each AP and \( x^{(l)}_k \) is given by Eq. (12). Furthermore, as the number of APs \( K_a \) grows sufficiently large, the system will experience only small variations (relative to the average) in the achievable data transmission rate, which is known as the channel hardening [19]. As a result, although the instantaneous CSI is not available at the mobile users, \( \mathbb{E}\left[ (g_{k,m})^T w_{k,m} \right] \) can be used to calculate the channel gain, where \( (\cdot)^T \) represents the transpose of a vector. Considering Rician wireless fading channels, we derive the received signal, denoted by \( y_m^{(l)} \), from the \( k \)th AP to the \( m \)th mobile user for transmitting the \( l \)th finite-blocklength data block as follows [11]:
\[ y_m^{(l)} = \sum_{k=1}^{K_u} \sqrt{P_d} \sqrt{\eta_{k,m}} \mathbb{E}\left[ (g_{k,m})^T w_{k,m} \right] s_m^{(l)} + \sqrt{P_d} \sum_{k=1}^{K_u} \sqrt{\eta_{k,m}} (g_{k,m})^T w_{k,m} - \mathbb{E}\left[ (g_{k,m})^T w_{k,m} \right] s_m^{(l)} \times \mathbb{E}\left[ (g_{k,m})^T w_{k,m} \right] s_m^{(l)} + \sum_{m' \neq m}^{K_u} \mathbb{E}\left[ (g_{k,m'})^T w_{k,m'} s_m^{(l)} + n_m^{(l)} \right] \] 
(16)

where \( s_m^{(l)} \) and \( s_m^{(l)} \) are the signals sent to mobile user \( m \) and mobile user \( m' \), respectively; \( \eta_{k,m} \) and \( \eta_{k,m'} \) are the power allocation parameters for transmitting from AP \( k \) to mobile user \( m \) and mobile user \( m' \), respectively; \( g_{k,m} \in \mathbb{C}^{1 \times N_t} \) represents the channel’s impulse response vector from the \( k \)th AP to mobile user \( m \); \( n_m^{(l)} \) is the AWGN with zero mean and unit variance; and \( D_{s_m}, B_{u_m}, \) and \( U_{l,m} \) represent the strength of the desired signal (DS), the beamforming gain uncertainty (BU), and the interference caused by the \( m' \)th mobile user (UI), respectively. Correspondingly, we can derive the signal-to-noise-plus-interference ratio (SINR), denoted by \( \gamma_m \), from the APs to mobile user \( m \) as follows:
\[ \gamma_m = \frac{\mathbb{E}\left[ \left\| D_{s_m} \right\|^2 \right]}{\mathbb{E}\left[ \left\| B_{u_m} \right\|^2 \right] + \sum_{m' = 1, m' \neq m}^{K_u} \mathbb{E}\left[ \left\| U_{l,m'} \right\|^2 \right]} + 1 \] 
(17)

III. STATISTICAL DELAY AND ERROR-RATE BOUNDED QoS PROVISIONING IN THE FINITE BLOCKLENGTH REGIME

In this section, we derive the Mellin transform over arrival and service processes by using SNC for our proposed CF m-MIMO schemes given the non-vanishing error probability.

A. \((n_d, M_m, \epsilon_m)\)-Code

Definition 1 \((n_d, M_m, \epsilon_m)\)-Code: We define a codebook consisting of \( M_m \) codewords, denoted by \( (c_1, \ldots, c_{M_m}) \), with length \( n_d \). We define a message set \( M_m = \{1, \ldots, M_m\} \) and a message \( W_m \in M_m \) which is uniformly distributed on \( M_m \). We define an \((n_d, M_m, \epsilon_m)\)-code \( (\epsilon_m \in [0, 1]) \) as follows:
- An encoder \( \Upsilon: \{1, \ldots, M_m\} \rightarrow \mathbb{C}^{N_t \times n_d} \) that maps the message \( W_m \in M_m \) to a codeword \( X_{n_d}^m \) with length \( n_d \).
- A decoder \( \Delta: \) a decoder \( \{D_{g_{k,m}}\} \rightarrow \mathbb{C}^{1 \times N_t} \times \mathbb{C}^{N_t \times n_d} \rightarrow \{1, \ldots, M_m\} \cup \{\epsilon\} \), where \( \epsilon \) represents the error event.

In [6], the accurate approximation of the maximum achievable data rate, denoted by \( R_m \) (bits per channel use), with error probability, denoted by \( \epsilon_m \) with \( 0 \leq \epsilon_m < 1 \), and coding blocklength, denoted by \( n_d \), for mobile user \( m \) in the finite blocklength regime can be determined as follows:
\[ R_m(n_d, \epsilon_m) \approx C(\gamma_m) - \frac{\sqrt{V(\gamma_m)}}{n_d} Q^{-1}(\epsilon_m) \] 
(18)

where \( Q^{-1}(\cdot) \) is the inverse of \( Q \)-function and \( C(\gamma_m) \) and \( V(\gamma_m) \) are the channel capacity and channel dispersion, respectively, which are given as follows [6]:
\[ \left\{ \begin{align*} C(\gamma_m) &= \log_2 (1 + \gamma_m) ; \\ V(\gamma_m) &= 1 - \frac{1}{(1 + \gamma_m)^2}. \end{align*} \right. \] 
(19)

B. Stochastic Network Calculus

Consider that each AP is equipped with a QoS-driven first-in-first-out (FIFO) buffer. We define \( a_m(l) \) as the source rate for transmitting the \( l \)th data block to mobile user \( m \) and \( s_m(l) \) as the instantaneous service rate over wireless channels for transmitting the \( l \)th data block to mobile user \( m \). Define \( A_m(l) \triangleq \sum_{j=0}^{l-1} a_m(j) \) as the accumulated source rate for transmitting \( l \) data blocks to mobile user \( m \) and \( S_m(l) \triangleq \sum_{j=0}^{l-1} s_m(j) \) as the accumulated service rate over wireless channels for transmitting \( l \) data blocks to mobile user \( m \). Define \( Q_m(l) \) as the dynamics of queueing process for transmitting \( l \) data blocks to mobile user \( m \), which is given as in the following equation:
\[ Q_m(l) = \max \{A_m(l) - S_m(l), 0\}. \] 
(20)
However, in practical scenarios, it is difficult to obtain the statistical characteristics of random arrival and service processes. As a result, by taking the exponential of arrival and service processes, we can transform the arrival and service processes, denoted by \( A_m(l) \) and \( S_m(l) \), respectively, in the bit domain into the exponential domain, i.e., signal-to-noise ratio (SNR)-domain [20] by using the exponential function given as follows:

\[
\begin{align*}
A_m(l) & \triangleq e^{A_m(l)}, \\
S_m(l) & \triangleq e^{S_m(l)}. 
\end{align*}
\]

(21)

Define the Mellin transform, denoted by \( M_X(\theta_m) \), of a non-negative random variable \( X \) as follows [21]:

\[
M_X(\theta_m) \triangleq \mathbb{E}[X^{(\theta_m-1)}] 
\]

(22)

where \( \theta_m > 0 \) is defined as the QoS exponent. Denoting \( d_{th} \) a target delay, we can define the kernel function \( K(\theta_m, d_{th}) \) as follows [20]:

\[
K(\theta_m, d_{th}) \triangleq \frac{M_{S_m}(1-\theta_m)^{d_{th}}}{1 - M_{A_m}(1 + \theta_m)M_{S_m}(1 - \theta_m)}, 
\]

(23)

if the following stability condition holds:

\[
M_{A_m}(1 + \theta_m)M_{S_m}(1 - \theta_m) < 1. 
\]

(24)

Correspondingly, an upper bound on the delay violation probability, denoted by \( p_m(d_{th}) \), can be obtained using the Mellin transform over the arrival and service processes \( A_m(l) \) and \( S_m(l) \) in the SNR-domain as follows:

\[
p_m(d_{th}) \leq \inf_{\theta_m > 0} \{ K(\theta_m, d_{th}) \}. 
\]

(25)

C. Statistical Delay and Error-Rate Bounded QoS Provisioning for CF m-MIMO in the Finite Blocklength Regime

Statistical delay-bounded QoS guarantees [22] [23] have been extensively studied to analyze queuing behavior for time-varying arrival and service processes. The PDF of SINR \( \gamma_m \) over Rician wireless fading channels is given as follows [24]:

\[
f_{\gamma_m}(\gamma_m) = 2(1 + \kappa)\gamma_m e^{-(1+\kappa)\gamma_m - \kappa} I_0 \left( 2\sqrt{\kappa (1 + \kappa) \gamma_m} \right) 
\]

(26)

where \( I_0[\cdot] \) is the 0th order modified Bessel function of the first kind.

1) Mellin Transform Over Arrival Process: Assume that the arrivals at all time slots are independent and i.i.d. for each mobile user \( m \), i.e., the accumulated source rate \( A_m(l) \) has i.i.d. increments, denoted by \( \alpha_m(l) \), or equivalently \( \alpha_m = a_m(l) \) due to \( a_m(l) \)'s being i.i.d. We can derive the Mellin transform over accumulated arrival process, denoted by \( M_{A_m}(\theta_m) \), as follows:

\[
M_{A_m}(\theta_m) = \mathbb{E} \left[ \left( \prod_{j=1}^{l} e^{\alpha_m(j)} \right)^{\theta_m-1} \right] = \left[ \mathbb{E} \left[ e^{\alpha_m(\theta_m-1)} \right] \right]^l, 
\]

(27)

where \( \alpha_m = e^{\alpha_m} \). Assume that the arrival process follows a Poisson distribution with average rate \( \lambda_m \). We can derive the Mellin transform of \( \alpha_m \) as follows:

\[
M_{\alpha_m}(\theta_m) = \sum_{i=1}^{\infty} e^{(\theta_m-1) \frac{(\lambda_m)^i}{i!} \lambda_m} = e^{\lambda_m (e^{\theta_m-1} - 1)}. 
\]

(28)

2) Mellin Transform Over Service Process: Equations (23) and (25) show that deriving the closed-form expression of Mellin transform over service process at mobile user \( m \) is important for analyzing the delay violation probability, which motivates the following theorem.

Theorem 1: Given the statistical delay and error-rate bounded QoS provisioning \( \{ \theta_m, \epsilon_m \} \), the Mellin transform over service process, denoted by \( M_{S_m}(1 - \theta_m) \), of mobile user \( m \) over Rician wireless fading channels in the high-end SNR region can be derived as follows:

\[
M_{S_m}(1 - \theta_m) = (1 - \epsilon_m) \left[ F_1(\gamma_0) + F_0(\gamma_0) \right] + \epsilon_m 
\]

(29)

where \( \gamma_0 \triangleq e^{\frac{\sqrt{\gamma_m}}{n_4} Q^{-1}(\epsilon_m)} \) and

\[
\begin{align*}
F_0(\gamma_0) & \triangleq 2 e^{-\kappa} \frac{\kappa}{\kappa + 1} \sum_{i=0}^{\infty} \frac{(i!2)^{\epsilon}}{(i!)^2} \gamma (i + 2, (1 + \kappa) \gamma_0); \\
F_1(\gamma_0) & \triangleq 2 e^{-\kappa} \left[ e^{-\sqrt{\gamma_m} Q^{-1}(\epsilon_m)} \right] \sum_{i=0}^{\infty} \frac{\kappa^{i+1}}{(i!2)^{\epsilon}} \times \Gamma \left( i + 2 - \frac{\theta_m n d}{\log 2} (1 + \kappa) \gamma_0 \right),
\end{align*}
\]

(30)

where \( \log(\cdot) \) represents \( \log_2(\cdot) \) and \( \gamma(a, b) \) and \( \Gamma(a, b) \) are the lower and upper incomplete Gamma functions, respectively.

Proof: Considering the decoding error at the receiver, we can derive the Mellin transform over service process as follows:

\[
M_{S_m}(1 - \theta_m) = \mathbb{E}_{\gamma_m} \left[ \epsilon_m + (1 - \epsilon_m) e^{-\theta_m n d R_m(n_d, \epsilon_m)} \right] \\
= \int_0^{\infty} \left[ \epsilon_m + (1 - \epsilon_m) e^{-\theta_m n d R_m(n_d, \epsilon_m)} \right] f_{\gamma_m}(\gamma_m) d\gamma_m 
\]

(31)

where \( \mathbb{E}_{\gamma_m}[\cdot] \) is the expectation operation with respect to \( \gamma_m \). We define

\[
f_m(n_d, \epsilon_m) \triangleq \frac{1 + \gamma_m}{\exp \left( \frac{\sqrt{\gamma_m}}{n_4} Q^{-1}(\epsilon_m) \right)}. 
\]

(32)

Given the decoding error probability \( \epsilon_m \), the data rate \( R_m(n_d, \epsilon_m) \) could become smaller than zero when the SINR is below a certain threshold \( \gamma_m \) [25]. As a result, the achievable data rate can be rewritten as follows:

\[
R_m(n_d, \epsilon_m) = \max \{ \log_2( f_m(n_d, \epsilon_m)), 0 \}. 
\]

(33)

Accordingly, we can obtain the following equation:

\[
M_{S_m}(1 - \theta_m) = (1 - \epsilon_m) \left\{ \int_0^{\infty} \left[ f_m(n_d, \epsilon_m) \right]^{-\frac{n_d}{\log 2}} \right\}. 
\]

(34)
under the stability condition $\mathcal{M}_{A_m}(1 + \theta_m)\mathcal{M}_{S_m}(1 - \theta_m) < 1$. As a result, we can derive the upper-bound on the delay violation probability by plugging Eq. (39) back into Eq. (25).

IV. DELAY ANALYSES FOR STATISTICAL DELAY AND ERROR-RATE BOUNDED QoS PROVISIONING IN THE FINITE BLOCKLENGTH REGIME

In the previous Section III, we have investigated an upper bound on the delay violation probability using the Mellin transform for a given decoding error probability $\epsilon_m$. In this section, assuming the decoding error probability is a function of $\{n_d, \gamma_m\}$, we derive the delay violation probability in terms of the average decoding error probability function over Rayleigh wireless fading channels ($\kappa = 0$).

A. Upper-Bound on the Average Decoding Error Probability Function for CF m-MIMO in the Finite Blocklength Regime

Consider the case where $\kappa = 0$, i.e., the Rayleigh fading channel model. We define [11]

$$c_{k,m} \triangleq \sqrt{n_p\phi_{p,k,m}} \sum_{m=1}^{K_s} \left\| \phi_{m}^{n_p} \left( \phi_{m}^{n_p} \right)^H \right\| + 1$$

Considering the massive access scenario with very large $K_s$, we can obtain the following equations by applying the Tchibessov’s theorem for our proposed CF m-MIMO schemes across Rayleigh fading channel model [11]:

$$\frac{1}{K_s} \sum_{k=1}^{K_s} \left( g_{k,m} \right)^T w_{k,m} = N_T \sqrt{n_p\phi_{p,k,m}} \sum_{k=1}^{K_s} \eta_{k,m} (c_{k,m}\beta_{k,m} \chi_{k})^2 \xrightarrow{K_s \to \infty} 0;$$

where the symbol $\xrightarrow{p}$ represents the convergence in probability as $K_s \to \infty$. The results given by Eq. (41) imply that as $K_s \to \infty$, we only need to consider the desired parts of the received signal $y_{m}^{(l)}$ and ignore the noise and interference in the asymptotic analysis. As a result, we can show that the SINR $\gamma_m$ follows the following distribution:

$$\gamma_m \sim \left( N_T^2 \right) \sum_{k=1}^{K_s} \eta_{k,m} (c_{k,m}\beta_{k,m} \chi_{k})^2 \xrightarrow{\xi_{k,m}} \sum_{k=1}^{K_s} \mathcal{E}(\xi_{k,m})$$

where $\mathcal{E}(\xi_{k,m})$ is the exponential distribution with its parameter equal to $\xi_{k,m}$, which is given as follows:

$$\xi_{k,m} = \frac{1}{2 \left( N_T^2 \right) \sum_{k=1}^{K_s} \eta_{k,m} (c_{k,m}\beta_{k,m})^2}$$

where $k \in \{1, \ldots, K_s\}$ and $m \in \{1, \ldots, K_s\}$. We can obtain the decoding error probability function, denoted by $\epsilon_m(n_d, \gamma_m)$, for mobile user $m$ as follows [6]:

$$\epsilon_m (n_d, \gamma_m) \approx Q \left( \frac{C(\gamma_m) - R_m}{\sqrt{V(\gamma_m)/n_d}} \right)$$

Therefore, by substituting Eqs. (36) and (38) back into Eq. (34), we can obtain the results in Eq. (29), which completes the proof of Theorem 1.

Correspondingly, substituting Eqs. (27), (28), and (29) back into Eq. (23), we can derive the closed-form expression of the steady-state kernel $K(\theta_m, \theta_{d_m})$ as follows:

$$K(\theta_m, \theta_{d_m}) = \frac{(1 - \epsilon_m) \left\{ [F_1(\gamma_m) + F_0(\gamma_m)] + \epsilon_m \right\} d\theta_m}{1 - e^{\lambda_m}d(\epsilon_m - 1)}$$
where \( Q(\cdot) \) is the \( Q \)-function, \( R_m \) (bits per channel use) is the achievable data rate, and \( C(\gamma_m) \) and \( V(\gamma_m) \) denote the channel capacity and channel dispersion, respectively, given in Eq. (19). Similar to Eq. (31), given the achievable finite-blocklength coding rate \( R_m \) and the decoding error probability function \( \epsilon_m(n_d, \gamma_m) \), we can derive the Mellin transform over service process \( S_m(l) \) at mobile user \( m \) as follows:

\[
\mathcal{M}S_m(1 - \theta_m) = \mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) + \left( 1 - \epsilon_m(n_d, \gamma_m) \right) e^{-\theta_m n_d R_m} \right] \\
= \mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] + \left( 1 - \mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \right) e^{-\theta_m n_d R_m}.
\]

Equation (45) shows that deriving the average decoding error probability function \( \mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \) is important to obtain the closed-form expression for Mellin transform over service process at mobile user \( m \), motivating the theorem that follows.

**Theorem 2:** Given the achievable finite-blocklength coding rate \( R_m \), the average decoding error probability function \( \mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \) for mobile user \( m \) over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime is determined as follows:

\[
\mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \approx 1 - \sum_{k=1}^{K_s} \left[ 1 - e^{-\xi_{k,m} \left( 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \right) } \right] + \left[ \frac{1}{2} + \theta_m \sqrt{n_d} \right] \times \left( e^{R_m - 1} \right) \\
- \sum_{k=1}^{K_s} \left[ \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \right] \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} + \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
+ \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} + \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
+ \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \left( 2^{R_m - 1} + \frac{1}{2 \theta_m \sqrt{n_d}} \right) } - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right)
\]

where \( Ei(x) = -\int_x^\infty \frac{e^{-t}}{t} \, dt \) represents the exponential integral function and \( \theta_m \triangleq \frac{1}{2 \pi 2^{R_m - 1}}. \)

**Proof:** To derive the average decoding error probability function, first we introduce an approximation of the \( Q \)-function as follows:

\[
Q \left( \frac{C(\gamma_m) - R_m}{\sqrt{V(\gamma_m)/n_d}} \right) \approx \Psi(\gamma_m)
\]

where the function \( \Psi(\gamma_m) \) is given as follows [27]:

\[
\Psi(\gamma_m) = \begin{cases} 
1, & \gamma_m \leq \zeta_{m,1}; \\
\frac{1}{2} - \theta_m \sqrt{n_d} \left( (\gamma_m - 2^{R_m - 1}) \right), & \zeta_{m,1} < \gamma_m < \zeta_{m,a}; \\
0, & \gamma_m \geq \zeta_{m,a},
\end{cases}
\]

(48)

where \( \theta_m \triangleq \frac{1}{2 \pi 2^{R_m - 1}}, \zeta_{m,1} \triangleq 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \), and \( \zeta_{m,a} \triangleq 2^{R_m - 1} + \frac{1}{2 \theta_m \sqrt{n_d}} \). Taking expectation over Eqs. (44) and (48), we can obtain the following equation:

\[
\mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \approx F_{\gamma_m} \left( 2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}} \right) + \left[ \frac{1}{2} + \theta_m \sqrt{n_d} \right] \left( e^{R_m - 1} \right) \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \int_{2^{R_m - 1} - \frac{1}{2 \theta_m \sqrt{n_d}}}^{2^{R_m - 1} + \frac{1}{2 \theta_m \sqrt{n_d}}} \gamma_m f_{\gamma_m}(\gamma_m) d\gamma_m
\]

(49)

where \( F_{\gamma_m}(\gamma_m) \) is the cumulative probability function (CDF) of SINR \( \gamma_m \). Using Eqs. (42) and (43), we can derive the CDF of SINR as follows:

\[
F_{\gamma_m}(\gamma_m) = 1 - \sum_{k=1}^{K_s} \left( 1 - e^{-\xi_{k,m} \gamma_m} \right).
\]

(50)

Plugging Eq. (50) back into Eq. (49), we obtain:

\[
\mathbb{E}_{\gamma_m} \left[ \epsilon_m(n_d, \gamma_m) \right] \approx 1 - \sum_{k=1}^{K_s} \left[ 1 - e^{-\xi_{k,m} \gamma_m} \right] + \left( \frac{1}{2} + \theta_m \sqrt{n_d} \right) \left( e^{R_m - 1} \right) \\
- \sum_{k=1}^{K_s} \left[ \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \right] \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
+ \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
- \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
+ \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right) \\
+ \left( \frac{1}{2 \theta_m \sqrt{n_d}} \right) \sum_{k=1}^{K_s} e^{-\xi_{k,m} \gamma_m} - \theta_m \sqrt{n_d} \left( e^{R_m - 1} \right)
\]

(51)

which, by employing the exponential integral function defined at the bottom of Eq. (46), leads to Eq. (46), completing the proof of Theorem 2.

In the high-end SNR regime, we have \( V(\gamma_m) = 1 - (1 + \gamma_m)^{-2} \rightarrow 1 \). Using Eq. (44), we define the following function:

\[
\Phi(n_d, \gamma_m) \triangleq \left[ C(\gamma_m) - R_m \right] \sqrt{n_d}.
\]

(52)
Considering the high-end SNR regime, we can derive the average decoding error probability function as detailed in the following lemma.

**Lemma 2:** The approximate average decoding error probability function \( \mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \) for mobile user \( m \) in the high-end SNR regime is determined as follows:

\[
\mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \approx \frac{\sqrt{\pi}}{2\sqrt{n_d}} \sum_{k=1}^{K_s} \xi_{k,m} \exp \left\{ \frac{(\xi_{k,m})^2}{2n_d} - \xi_{k,m} \nu_m \right\} \times \left[ 1 - \text{erf} \left( \frac{\sqrt{\nu_m}}{\sqrt{2}} \right) \right]
\]

(53)

where \( \nu_m \triangleq 2R_m - 1 \).

**Proof:** Using the Chernoff bound, we have \( Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}} \) if \( x \geq 0 \). We can derive an upper bound on the average decoding error probability function \( \mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \) for mobile user \( m \) in the high-end SNR regime as follows:

\[
\mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] = \int_0^\infty \mathbb{E}_{\gamma_m}[\Phi(n_d, \gamma_m)] f_{\gamma_m}(\gamma_m)d\gamma_m \\
\leq \frac{1}{2} \int_0^\infty \mathbb{E}_{\gamma_m}[e^{-\frac{1}{2}(\gamma_m - \nu_m)^2}] f_{\gamma_m}(\gamma_m)d\gamma_m
\]

(54)

where \( f_{\gamma_m}(\gamma_m) \) is the PDF of the SINR \( \gamma_m \) over Rayleigh wireless fading channels, which is given as follows:

\[
f_{\gamma_m}(\gamma_m) = \sum_{k=1}^{K_s} \xi_{k,m} e^{-\xi_{k,m} \gamma_m}.
\]

(55)

Then, using Eq. (19), we can obtain:

\[
\mathbb{E}_{\gamma_m}[e^{-\frac{1}{2}(\gamma_m - \nu_m)^2}] = \mathbb{E}_{\gamma_m}[e^{-\frac{1}{2}(\gamma_m - \nu_m)^2}] \\
= \int_0^\infty e^{-\frac{1}{2}(\gamma_m - \nu_m)^2} f_{\gamma_m}(\gamma_m)d\gamma_m
\]

(56)

Since we can easily show that \( \Phi(n_d, \gamma_m) \) is concave in \( \gamma_m \), as a result, we can derive a lower bound on the function \( \text{delay and error-rate bounded QoS provisioning} \)

\[
\mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \\
= \frac{\sqrt{\pi}}{2\sqrt{n_d}} \sum_{k=1}^{K_s} \xi_{k,m} \exp \left\{ \frac{(\xi_{k,m})^2}{2n_d} - \xi_{k,m} \nu_m \right\} \times \left[ 1 - \text{erf} \left( \frac{\sqrt{\nu_m}}{\sqrt{2}} \right) \right]
\]

(57)

Letting \( t_m \triangleq \sqrt{\nu_m} (\gamma_m - \nu_m) \), we have

\[
\mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \\
\geq \sqrt{\frac{2}{n_d}} \sum_{k=1}^{K_s} \xi_{k,m} e^{-\xi_{k,m} \nu_m} \int_{-\frac{\sqrt{T_m}}{\sqrt{\nu_m}}}^{\frac{\sqrt{T_m}}{\sqrt{\nu_m}}} e^{-(t_m)^2} e^{-\frac{\sqrt{T_m}}{\sqrt{\nu_m}}} dt_m
\]

\[
= \sqrt{\frac{2}{n_d}} \sum_{k=1}^{K_s} \xi_{k,m} e^{-\xi_{k,m} \nu_m} \int_{-\frac{\sqrt{T_m}}{\sqrt{\nu_m}}}^{\frac{\sqrt{T_m}}{\sqrt{\nu_m}}} e^{-(t_m)^2} dt_m
\]

(58)

Letting \( t_m \triangleq t_m + \frac{\xi_{k,m}}{\sqrt{n_d}} \), we can derive a lower bound on the function \( \mathbb{E}_{\gamma_m}[e^{-\frac{1}{2}(\gamma_m - \nu_m)^2}] \) as follows:

\[
\mathbb{E}_{\gamma_m}[e^{-\frac{1}{2}(\gamma_m - \nu_m)^2}] \\
\geq \sqrt{\frac{2}{n_d}} \sum_{k=1}^{K_s} \xi_{k,m} \exp \left\{ \frac{(\xi_{k,m})^2}{2n_d} - \xi_{k,m} \nu_m \right\} \times \left[ 1 - \text{erf} \left( \frac{\sqrt{\nu_m}}{\sqrt{2}} \right) \right]
\]

(59)

where \( \text{erf}(x) \triangleq \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \) is the error function. As a result, by plugging Eq. (59) back into Eq. (54), we can obtain Eq. (53), which completes the proof of Lemma 2.

**B. Performance Analyses and Rate Adaptation for Statistical Delay and Error-Rate Bounded QoS Provisioning**

Substituting Eqs. (28), (29), and (46) back into Eq. (23), we can derive the closed-form expression for the kernel function \( \mathcal{K}(\theta, d_m) \). Correspondingly, we formulate the delay violation probability minimization problem as follows:

\[
P_1 : R_m = \arg \min_{R_m} \{ p_m(d_m) \} = \arg \min_{R_m} \{ \mathcal{K}_m(\theta, d_m) \}.
\]

(60)

Using Eq. (23), we can convert \( P_1 \) into an equivalent minimization problem \( P_2 \) as in the following equation:

\[
P_2 : R_m = \arg \min_{R_m} \{ \mathcal{M}_S_m(1 - \theta,m) \} \\
= \arg \min_{R_m} \left\{ \mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] + \left[ 1 - e^{-\theta_m n_d R_m} \right] \times \mathbb{E}_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \right\}.
\]

(61)

The monotonicity of decoding error probability function \( \epsilon_m(n_d, \gamma_m) \) is an important role in analyzing the convexity of \( P_2 \) given in Eq. (60) as detailed in the following lemma.

**Lemma 3:** The decoding error probability function \( \epsilon_m(n_d, \gamma_m) \) is a monotonically increasing function with respect to the achievable data rate \( R_m \) for our proposed CF m-MIMO modeling schemes.

**Proof:** To prove the monotonicity of the decoding error probability function \( \epsilon_m(n_d, \gamma_m) \), using Eq. (44), we can take the first-order derivative of \( \epsilon_m(n_d, \gamma_m) \) with respect to \( R_m \) as in the following equation:

\[
\frac{\partial \epsilon_m(n_d, \gamma_m)}{\partial R_m} = -\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(\gamma_m - \nu_m)^2} \left[ \frac{\partial \Phi(n_d, \gamma_m)}{\partial R_m} \right]
\]

(62)
where \( \Phi (n_d, \gamma_m) \triangleq \frac{C(\gamma_m) - R_m}{\sqrt{V(\gamma_m)/n_d}} \) due to Eq. (44) and thus its first-order derivative yields the following equations:

\[
\frac{\partial \Phi (n_d, \gamma_m)}{\partial R_m} = \frac{-\sqrt{n_d}}{\sqrt{1 - \frac{(1 + \gamma_m)^2}{2}}} < 0. \tag{63}
\]

Therefore, we have \( \frac{\partial \epsilon_m(n_d, \gamma_m)}{\partial R_m} > 0 \), which implies that the decoding error probability function \( \epsilon_m(n_d, \gamma_m) \) is a monotonically increasing function of \( R_m \), completing the proof. \( \square \)

Lemma 3 and its proof can help further investigate the convexity of \( \epsilon_m(n_d, \gamma_m) \) as shown in the following lemma.

**Lemma 4.** The block error probability function \( \epsilon_m(n_d, \gamma_m) \) is convex with respect to the achievable data rate \( R_m \) for each mobile user \( m \).

**Proof:** Applying Eq. (63), we can derive the second-order derivative of the function \( \Phi (n_d, \gamma_m) \) with respect to \( R_m \) as follows:

\[
\frac{\partial^2 \Phi (n_d, \gamma_m)}{\partial R_m^2} = 0. \tag{64}
\]

Using Eqs. (44), (63), and (64) and the fact that \( \Phi (n_d, \gamma_m) > 0 \) due to \( C(\gamma_m) > R_m \), we obtain the following equations:

\[
\frac{\partial^2 \epsilon_m(n_d, \gamma_m)}{\partial R_m^2} = \frac{1}{\sqrt{2\pi}} e^{-\frac{\epsilon^2(n_d, \gamma_m)}{2}} \left\{ \Phi(n_d, \gamma_m) \left[ \frac{\partial \Phi (n_d, \gamma_m)}{\partial R_m} \right] \right\}^2 - \frac{\partial^2 \Phi (n_d, \gamma_m)}{\partial R_m^2} > 0 \tag{65}
\]

which implies that \( \epsilon_m(n_d, \gamma_m) \) is a convex function with respect to \( R_m \). Therefore, we complete the proof. \( \square \)

Combining and extending Theorem 2, Lemma 3, and Lemma 4 yield the following theorem.

**Theorem 3.** If the statistical delay and error-rate bounded QoS constraints are characterized by \( \{\theta_m, \epsilon_m(n_d, \gamma_m)\} \) with \( \theta_m > 0 \), then the following claims hold for each mobile user \( m \).

**Claim 1.** The delay upper-bound violation probability minimization problem \( P_2 \) given by Eq. (61) is strictly convex with respect to the achievable data rate \( R_m \) for our proposed CM-MIMO modeling schemes.

**Claim 2.** Our obtained Mellin transform function \( M_{S_m}(1 - \theta_m) \) over service process satisfies the following condition:

\[
\frac{\partial^2 M_{S_m}(1 - \theta_m)}{\partial R_m^2} > 0. \tag{66}
\]

**Claim 3.** The unique optimal rate adaptation policy, denoted by \( R_m^{\text{opt}} \), for each mobile user \( m \) in the high-end SNR region is given by the following equation:

\[
R_m^{\text{opt}} \approx \log_2 \left\{ \th M \right\} \th \sum_{k=1}^{K_s} \xi_{k,m} \left( \frac{\log 2}{\theta_m n_d} \right) - W \left( \frac{(log 2) \sum_{k=1}^{K_s} \xi_{k,m}}{\theta_m n_d} \right) \left( 1 - \frac{\theta_m \sqrt{2 \pi n_d}}{2} \right)^{\frac{\log 2}{2}}. \tag{72}
\]

where \( W(\cdot) \) is the Lambert \( W \) function.

**Proof:** Applying Eq. (66), we can derive the following expression for each mobile user \( m \):

\[
\frac{\partial^2 M_{S_m}(1 - \theta_m)}{\partial R_m^2} = \frac{\partial^2 M_{S_m}(1 - \theta_m)}{\partial R_m^2} \left[ \frac{\partial \epsilon_m(n_d, \gamma_m)}{\partial R_m} \right]^2 + \frac{\partial \epsilon_m(n_d, \gamma_m)}{\partial R_m} \frac{\partial^2 \epsilon_m(n_d, \gamma_m)}{\partial R_m^2}. \tag{68}
\]

To analyze the convexity of the Mellin transform over service process \( M_{S_m}(1 - \theta_m) \), first we need to investigate the following equations when \( \theta_m > 0 \):

\[
\frac{\partial M_{S_m}(1 - \theta_m)}{\partial \epsilon_m(n_d, \gamma_m)} = 1 - e^{-\theta_m n_d R_m} > 0; \tag{69}
\]

\[
\frac{\partial^2 M_{S_m}(1 - \theta_m)}{\partial \epsilon_m(n_d, \gamma_m)^2} = 0. \tag{69}
\]

Then, using Eq. (65) and (69), we can show that \( \frac{\partial^2 M_{S_m}(1 - \theta_m)}{\partial \epsilon_m(n_d, \gamma_m)^2} > 0 \), completing the proof of Claim 2 in Theorem 3.

**Claim 1.** Applying the sufficient conditions for convexity, we can prove that the Mellin transform function \( M_{S_m}(1 - \theta_m) \) is strictly convex with respect to the achievable data rate \( R_m \) for \( \theta_m > 0 \), completing the proof of Claim 1 in Theorem 3.

**Claim 3.** Due to the property of strict convexity and uniqueness of optimal solutions [28], there exists the unique optimal rate adaptation policy \( R_m^{\text{opt}} \) that minimizes problem \( P_2 \) given by Eq. (61) for each mobile user \( m \) when \( \theta_m > 0 \). Taking the first-order derivative of the Mellin transform function \( M_{S_m} \) with respect to \( R_m \), we can obtain the following equation:

\[
\frac{\partial M_{S_m}(1 - \theta_m)}{\partial R_m} = \frac{\partial E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)]}{\partial R_m} - \frac{\partial E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)]}{\partial R_m} \times e^{-\theta_m n_d R_m} - \left\{ 1 - E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \right\} \times \theta_m n_d e^{-\theta_m n_d R_m}. \tag{70}
\]

Setting \( \frac{\partial M_{S_m}(1 - \theta_m)}{\partial R_m} = 0 \), we have

\[
\frac{\partial E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)]}{\partial R_m} \left( 1 - e^{-\theta_m n_d R_m} - \left\{ 1 - E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)] \right\} \right) \times \theta_m n_d e^{-\theta_m n_d R_m} = 0. \tag{71}
\]

We can rewrite Eq. (71) as follows:

\[
e_{\gamma_m(n_d, \gamma_m)} - 1 = \frac{\theta_m n_d (1 - E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)])}{\partial E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)]} \frac{\partial E_{\gamma_m}[\epsilon_m(n_d, \gamma_m)]}{\partial R_m}. \tag{72}
\]
Using Eqs. (62) and (63), we can obtain the following equation:

\[ R_m = \frac{1}{\theta_m n_d} \log \left\{ 1 + \frac{\theta_m n_d \left( 1 - E_{\gamma_m} \left( \mu_m (n_d, \gamma_m) \right) \right)}{\theta_m n_d \left( 1 - E_{\gamma_m} \left( \mu_m (n_d, \gamma_m) \right) \right)} \right\} \]

(73)

where \( E_{\gamma_m} \left( \mu_m (n_d, \gamma_m) \right) \) is given by Eq. (46). In the high-end SNR regime, we have \( V (\gamma_m) = 1 - (1 + \gamma_m)^{-2} \rightarrow 1 \). Using the Chernoff bound, we have \( Q(x) \leq \frac{1}{2} e^{-\frac{x^2}{2}} \) when \( x \geq 0 \).

Applying Eq. (73), we can obtain the following equation:

\[ R_m = \frac{1}{\theta_m n_d} \log \left\{ 1 + \frac{\theta_m n_d \left( 1 - E_{\gamma_m} \left( \mu_m (n_d, \gamma_m) \right) \right)}{E_{\gamma_m} [e^{-\phi(n_d, \gamma_m)}]} \right\} \]

(74)

Then, plugging Eq. (59) back into Eq. (74), we have:

\[ R_m \approx \frac{1}{\theta_m n_d} \log \left( 1 + \theta_m \sqrt{2\pi n_d} \left( -1 + \frac{1}{2} + \frac{\sqrt{2n_d}}{\sqrt{\pi}} \right) \right) \]

\[ \times \left[ \sum_{k=1}^{K} \xi_{k,m} \exp \left( \frac{(\xi_{k,m})^2}{2n_d} - \xi_{k,m} \nu_m \right) \right] \]

\[ \times \left\{ 1 - \text{erf} \left( \frac{\xi_{k,m}}{\sqrt{2n_d}} \right) \right\} \]

(75)

Since the error function \( -1 \leq \text{erf}(x) \leq 1 \), we can remove the error function from Eq. (75) and obtain the following equation:

\[ e^{\theta_m n_d R_m} \approx 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} + 2\theta_m n_d \sum_{k=1}^{K} \xi_{k,m} \]

\[ \times \left\{ \exp \left( \frac{(\xi_{k,m})^2}{2n_d} - \xi_{k,m} \nu_m \right) \right\} \]

(76)

Substituting \( \nu_m = 2R_m - 1 \) back into Eq. (76), we have

\[ e^{\theta_m n_d R_m} - \sum_{k=1}^{K} 2\theta_m n_d e^{\xi_{k,m} 2R_m} \approx 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} \]

(77)

where \( \xi_{k,m} \) is given by Eq. (43). Let \( z_m = \frac{\theta_m n_d}{\log 2} \).

We have

\[ \left( K \sum_{k=1}^{K} \xi_{k,m} \right) - \frac{\theta_m n_d}{\log 2} \rightarrow 2\theta_m n_d e^{z_m} \sum_{k=1}^{K} \left( \frac{\xi_{k,m}}{2n_d} + \xi_{k,m} \right) \]

\[ \times \exp \left( \frac{(\xi_{k,m})^2}{2n_d} + \xi_{k,m} \right) \approx 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} \]

(78)

Then, solving for the \( \theta_m n_d \) th root on both sides of Eq. (78), we can obtain:

\[ \left( \sum_{k=1}^{K} \xi_{k,m} \right) ^{-1} \approx z_m - e^{-z_m - \frac{\log 2}{\theta_m n_d}} \]

(79)

Let \( z_m = \frac{\theta_m n_d}{\log 2} \). After some algebra manipulations, we have

\[ \frac{(\log 2) K \sum_{k=1}^{K} \xi_{k,m}}{\theta_m n_d} = \frac{2\theta_m n_d}{\theta_m n_d} \sum_{k=1}^{K} \xi_{k,m} \exp \left( \frac{(\xi_{k,m})^2}{2n_d} + \xi_{k,m} \right) \]

\[ \approx z_m - \frac{(\log 2) K \sum_{k=1}^{K} \xi_{k,m}}{\theta_m n_d} \left( 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} \right) \]

(80)

Then, multiplying the following expression on both sides of Eq. (80):

\[ e^{-z_m - \frac{(\log 2) K \sum_{k=1}^{K} \xi_{k,m}}{\theta_m n_d}} \left( 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} \right) \]

(81)

we can obtain:

\[ \frac{(\log 2) K \sum_{k=1}^{K} \xi_{k,m}}{\theta_m n_d} = \frac{2\theta_m n_d}{\theta_m n_d} \sum_{k=1}^{K} \xi_{k,m} \exp \left( \frac{(\xi_{k,m})^2}{2n_d} + \xi_{k,m} \right) \]

\[ \times \exp \left( \frac{(\log 2) K \sum_{k=1}^{K} \xi_{k,m}}{\theta_m n_d} \left( 1 - \frac{\theta_m \sqrt{2\pi n_d}}{2} \right) \right) \]

(82)
Therefore, using the Lambert $W$ function [29], we can derive the approximate optimal rate adaptation policy $R_{m}^{\text{opt}}$ for our proposed FBC based CF m-MIMO schemes in the high-end SNR region as follows:

\[ R_{m}^{\text{opt}} \approx \log_{2} \left\{ \frac{\theta_{m} n_{d}}{(\log 2) \sum_{k=1}^{K_{m}} \xi_{k,m}} \left[ (\log 2) \frac{\sum_{k=1}^{K_{m}} \xi_{k,m}}{\theta_{m} n_{d}} \right]^{-W} \times \left( 1 - \frac{\theta_{m} \sqrt{2\pi n_{d} \xi_{k,m}}}{2} \right)^{-W} \right\}. \]  

Thus, we complete the proof of Claim 3 in Theorem 3. \hfill \square

Remarks on Theorem 3: Claim 1 guarantees the existence of the optimal solution to the optimization problem $P_{2}$ given by Eq. (61) for our proposed CF m-MIMO modeling schemes when $\theta_{m} > 0$. Claim 2 shows that our obtained Mellin transform function over service process can characterize the convexity of $P_{2}$. Claim 3 derives the closed-form expression of the optimal rate adaptation policy $R_{m}^{\text{opt}}$ as the function of \{\theta_{m}, \epsilon_{m} (n_{d}, \gamma_{m})\} for each mobile user $m$ in the high-end SNR region, which plays the important roles in the system designs and performance analyses for statistical delay and error-rate bounded QoS provisioning over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime.

C. Maximizing Effective Capacity for Statistical Delay and Error-Rate Bounded QoS Constraints in the Finite Blocklength Regime

The proposed upper bound on the delay violation probability $p_{m}(d_{th})$ using SNC characterizes the small values of the target delay $d_{th}$. For analyzing fairly long delays, i.e., the tail of delay distribution, researchers have proposed the concept of effective capacity to approximate the delay violation probability $p_{m}(d_{th})$. The effective capacity [30] is defined as the maximum constant arrival rate for a given service process subject to statistical delay-bounded QoS constraints. We can derive the effective capacity, denoted by $EC_{m}(\theta_{m})$, in terms of the accumulated service process $S_{m}(l)$ for mobile user $m$ in the bit domain as in the following equation:

\[ EC_{m}(\theta_{m}) \triangleq \lim_{l \to \infty} \frac{1}{\theta_{m}} \log \left( \mathbb{E}_{S_{m}} \left[ e^{-\theta S_{m}(l)} \right] \right). \]  

On the other hand, considering the SNC, we can redefine the effective capacity, denoted by $EC_{m}(\theta_{m})$, using the Mellin transform over service process in the SNR-domain as follows:

\[ EC_{m}(\theta_{m}) = -\frac{1}{\theta_{m}} \log (\mathcal{M}_{S_{m}}(1 - \theta_{m})). \]  

Accordingly, we can formulate the optimization problem $P_{3}$ for statistical delay and error-rate bounded QoS provisioning \{\theta_{m}, \epsilon_{m} (n_{d}, \gamma_{m})\} to maximize the downlink effective capacity $EC_{m}(\theta_{m})$ at mobile user $m$ for our proposed CF m-MIMO schemes in the finite blocklength regime as follows:

\[ P_{3} : \arg \max_{\theta_{m}} \{ EC_{m}(\theta_{m}) \}. \]

Since $\log(\cdot)$ is monotonically increasing, the above maximization problem $P_{3}$ can be converted into an equivalent minimization problem $P_{4}$ as follows:

\[ P_{4} : \arg \min_{\theta_{m}} \{ \mathcal{M}_{S_{m}}(1 - \theta_{m}) \}. \]

which is equivalent to the minimization problem $P_{2}$ given in Eq. (61). Consequently, we can show that our derived optimal rate adaptation policy $R_{m}^{\text{opt}}$ given in Eq. (67) also maximizes the effective capacity $EC_{m}(\theta_{m})$ given in Eq. (85) for mobile user $m$ in the finite blocklength regime considering the high-end SNR region. Therefore, using Eq. (45), (53), and (67), we can derive the maximum effective capacity, denoted by $EC_{m}^{\max}(\theta_{m})$, for statistical delay and error-rate bounded QoS provisioning in supporting mMTC over 6G CF m-MIMO and FBC mobile wireless networks in the high-end SNR region as follows:

\[ EC_{m}^{\max}(\theta_{m}) \approx -\frac{1}{\theta_{m}} \log \left( \frac{\sqrt{2\pi n_{d}}}{2} \sum_{k=1}^{K_{m}} \xi_{k,m} \exp \left( \frac{(\xi_{k,m})^{2}}{2n_{d}} - \xi_{k,m} \nu_{m} \right) \right) \times \left[ 1 - \text{erf} \left( \frac{\xi_{k,m}}{\sqrt{2n_{d}}} - \sqrt{n_{d}} \nu_{m} \right) \right] \right\} \times \frac{\theta_{m} n_{d}}{(\log 2) \sum_{k=1}^{K_{m}} \xi_{k,m}} \left[ (\log 2) \frac{\sum_{k=1}^{K_{m}} \xi_{k,m}}{\theta_{m} n_{d}} \right]^{-W} \times \left( 1 - \frac{\theta_{m} \sqrt{2\pi n_{d} \xi_{k,m}}}{2} \right)^{-W} \right\}. \]
\[ \times \exp \left[ \frac{(\log 2)}{\theta_T n_d} \left( \sum_{k=1}^{K_a} \xi_{k,m} \right) \right] \times \left( 1 - \frac{\theta_T \sqrt{2 \pi n_d}}{2} \right)^{\frac{(\log 2)}{\theta_T n_d}} \]. \quad (88) \]

V. PERFORMANCE EVALUATIONS

We use MATLAB-based simulations to validate and evaluate our proposed CF m-MIMO based schemes for statistical delay and error-rate bounded QoS provisioning in supporting mURLLC in the finite blocklength regime. Throughout our simulations, we set the number of APs \( K_a \in [50, 900] \), the number of mobile users \( K_d \in [10, 700] \), the number of transmit antennas \( N_T \in [2, 10] \), the uplink pilot transmit power \( P_p \) from [1, 10] Watt for each mobile user, the average downlink transmit power \( P_d \) from [1, 40] Watt for each mobile user, and the Rician factor \( \kappa \) from [0, 30].

We set the number of Rician factor \( \kappa = 4 \), the number of uplink channel uses \( n_p = 100 \), the number of transmit antennas \( N_T = 10 \), and the decoding error probability \( \epsilon_m = 10^{-6} \). Compared with the classical least-square (LS) channel estimator, Fig. 2 plots the achievable data transmission rate with varying numbers of APs \( K_a \) for our proposed 6G CF m-MIMO mobile wireless networks over Rician wireless fading channels in the finite blocklength regime. We can observe from Fig. 2 that the achievable data transmission rate increases with the number of APs. It is shown in Fig. 2 that the MMSE channel estimator performs better than the LS estimator over Rician wireless fading channels in terms of the achievable data transmission rate. Fig. 2 also shows that the gap between the MMSE estimator and LS estimator increases with \( K_a \), which is because of the channel hardening effect.

Setting the number of transmit antennas \( N_T = 10 \) and the decoding error probability \( \epsilon_m = 10^{-6} \). Fig. 3 depicts the achievable data transmission rate with different Rician factors \( \kappa \) for our proposed 6G CF m-MIMO mobile wireless networks in the finite blocklength regime. We can observe from Fig. 3 that the achievable data rate increases as the Rician factor \( \kappa \) increases. Traditionally, the CSI estimation is not good enough when \( K_a \) is small, which leads to a low data rate. For our proposed CF m-MIMO scheme, the channel estimation quality can be significantly improved with large number of APs \( K_a \).

Now we set the number of APs \( K_a = 100 \), the number of downlink channel uses \( n_d = 800 \), the average downlink transmit power \( P_d = 20 \) Watt for each AP, and the Rician factor \( \kappa = 2 \). Fig. 4 plots the CDFs of the downlink data transmission rate per user for our proposed 6G CF m-MIMO schemes in the finite blocklength regime. As shown in Fig. 4, the downlink data transmission rate per user increases with the number of transmit antennas \( N_T \). In addition, Fig. 4 shows that a higher multiplexing order \( P_m \) is more beneficial for our proposed CF m-MIMO schemes with larger antenna arrays.

We set the number of APs \( K_a = 100 \), the number of transmit antennas \( N_T = 10 \), the multiplexing order \( P_m = 2 \), and decoding error probability \( \epsilon_m = 10^{-6} \). Compared with the traditional Shannon’s theorem which requires infinite blocklength, Fig. 5 plots the delay violation probability \( p_m(d_{th}) \) with different target delays \( d_{th} \) over Rician wireless fading channels for our proposed CF m-MIMO scheme in the finite blocklength regime. It is shown in Fig. 5 that the delay violation probability \( p_m(d_{th}) \) decreases as the target delay \( d_{th} \) increases. Fig. 5 also shows the delay violation probability increases with the increased average data arrival rate \( \lambda_m \).
Fig. 5. The delay violation probability $p_m(d_h)$ vs. target delay $d_h$ for our proposed CF m-MIMO scheme in the finite blocklength regime.

Fig. 6. The delay (ms) vs. average arrival rate $\lambda_m$ for our proposed CF m-MIMO scheme in the finite blocklength regime. We set the target delay $d_h = 5$ ms, Rician factor $\kappa = 10$, the number of downlink channel uses $n_d = 800$, the average downlink transmit power $P_d = 20$ Watt for AP, and decoding error probability $\epsilon_m = 10^{-6}$. Fig. 6 depicts the delay in millisecond (ms) with varying average arrival rates $\lambda_m$ over Rician wireless fading channels for our proposed CF m-MIMO scheme in the finite blocklength regime, which implies the potential to support massive number of mobile users. We can observe from Fig. 6 that the queuing delay increases as the average arrival rate $\lambda_m$ increases. Fig. 6 also shows that the analytical results provide a reasonable upper bound for the actual delay as obtained from simulations.

We set the number of APs $K_a = 100$, the number of transmit antennas $N_T = 10$, the multiplexing order $P_m = 2$, Rician factor $\kappa = 0$, and the average downlink transmit power $P_d = 20$ Watt for each AP. Using Eq. (46), Fig. 7 depicts the average decoding error probability function $E_{\gamma_m} [\epsilon_m (n_d, \gamma_m)]$ with different achievable finite-blocklength coding rates $R_m$ for our proposed CF m-MIMO scheme across Rayleigh wireless fading channels in the finite blocklength regime. We can observe from Fig. 7 that the average decoding error probability function $E_{\gamma_m} [\epsilon_m (n_d, \gamma_m)]$ increases as the achievable finite-blocklength coding rate $R_m$ increases. Fig. 7 also shows that the gap between the simulated average decoding error function and the approximate average decoding error function is reasonably small.

We set the number of APs $K_a = 100$, the number of transmit antennas $N_T = 10$, the multiplexing order $P_m = 2$, Rician factor $\kappa = 0$, the number of downlink channel uses $n_d = 800$, the average downlink transmit power $P_d = 20$ Watt, and the average downlink transmit power $P_d = 20$ Watt. Compared with the scheme without the optimal rate adaptation (RA), Fig. 8 depicts the delay violation probability $p_m(d_h)$ with different numbers of APs $K_a$ for our proposed CF m-MIMO scheme in the finite blocklength regime. We can observe from Fig. 8 that the delay violation probability $p_m(d_h)$ decreases as the numbers of AP $K_a$ increases. Fig. 8 also shows that our proposed schemes with optimal RA outperform the schemes without applying the optimal RA in terms of the delay violation probability over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime.

Given different numbers of mobile users $K_u$, Fig. 9 depicts the block error probability function $\epsilon_m (n_d, \gamma_m)$ with varying blocklengths $n_d$ for our proposed CF m-MIMO scheme in the finite blocklength regime. We can observe from Fig. 9 that the performance degradation in terms of block error probability function $\epsilon_m (n_d, \gamma_m)$ with the increasing number of mobile users $K_u$ is mild, implying the remarkable potential as well as
the strong and robust scalability in supporting massive access by vast mobile devices over our proposed 6G CF m-MIMO mobile wireless networks.

Setting the number of transmit antennas $N_T = 2$, Rician factor $\kappa = 0$, and the average downlink transmit power $P_d = 20$ Watt, Fig. 10 plots the data transmission rate per user with different numbers of mobile users $K_u$ for our proposed CF m-MIMO scheme in the finite blocklength regime. We can observe from Fig. 10 that the data transmission rate per user decreases as the number of mobile users $K_u$ increases and will finally converge to a certain value, which implies the potential to support massive number of mobile users. Fig. 10 also shows that loose QoS constraint ($\theta_m = 10^{-3}$) and stringent QoS constraint ($\theta_m = 0.5$) set the upper bound and lower bound on the data transmission rate per user, receptively.

Figure 11 plots the maximum effective capacity $EC_{max}^{\theta_m}(n_d)$ with different blocklengths $n_d$ and QoS exponents $\theta_m$ for our proposed CF m-MIMO scheme in the finite blocklength regime. We can observe from Fig. 11 that the maximum effective capacity $EC_{max}^{\theta_m}(n_d)$ decreases as the QoS exponent $\theta_m$ increases. Fig. 11 also shows that the maximum effective capacity $EC_{max}^{\theta_m}(\theta_m)$ is an increasing function of the blocklength $n_d$ over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime.

VI. CONCLUSION

We have developed an analytical model to quantitatively characterize the performance for statistical delay and error-rate bounded QoS provisioning in supporting mURLLC over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime. In particular, we have developed CF m-MIMO based system models. Then, we have applied the Mellin transform to model and characterize both arrival and service processes, derived closed-form expressions for the delay violation probability function, and formulated and solved the delay violation probability minimization problem for our proposed CF m-MIMO modeling schemes in the finite blocklength regime. Furthermore, applying our developed system modeling techniques, we have derived a closed-form solution for the optimal rate adaptation policy, which plays an important role in system design and performance analyses for statistical delay and error-rate bounded QoS provisioning over 6G CF m-MIMO mobile wireless networks in the finite blocklength regime. We also have conducted a set of simulations to validate and evaluate our proposed CF m-MIMO schemes and show that our proposed schemes outperform the other existing schemes for statistical delay and error-rate bounded QoS provisioning in the finite blocklength regime.

REFERENCES


Jingqing Wang received the B.S. degree in electronics and information engineering from Northwestern Polytechnical University, Xi’an, China. She is currently pursuing the Ph.D. degree with Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, TX, USA, under the supervision of Prof. Xi Zhang. Her research interests include big data-based 5G wireless networks technologies, statistical QoS provisioning, and cognitive radio networks. She won the Best Paper Award from the IEEE GLOBECOM in 2014, the Hagler Institute for Advanced Study Heep Graduate Fellowship Award from Texas A&M University in 2018, and Dr. R.K. Pandey and Christa U. Pandey’84 Fellowship, Texas A&M University, USA, 2020–2021.

H. Vincent Poor (Life Fellow, IEEE) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana-Champaign. Since 1990, he has been on the faculty at Princeton, where he is currently the Michael Henry Strater University Professor of Electrical Engineering. Among his publications in these areas is the recent book, Multiple Access Techniques for 5G Wireless Networks and Beyond (Springer, 2019). From 2006 to 2016, he served as the Dean of the School of Engineering and Applied Science, Princeton University. He has also held visiting appointments at several other universities, including most recently at Berkeley and Cambridge. His research interests include information theory, signal processing, machine learning and their applications in wireless networks, and energy systems and their related fields. Among his publications in these areas is the recent book, Multiple Access Techniques for 5G Wireless Networks and Beyond (Springer, 2019).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Chinese Academy of Sciences, the Royal Society, and other national and international academies. His recent recognition of his work includes the 2017 IEEE Alexander Graham Bell Medal.