

# Effective Capacity of Non-Orthogonal Multiple Access with Finite Blocklength for Low-Latency Communications

Zina Mohamed, *Member, IEEE*, Muhammad Amjad, *Member, IEEE*,  
Leila Musavian, *Member, IEEE*, and Sonia Aïssa, *Fellow, IEEE*

**Abstract**—In this paper, we focus on investigating the link-layer rate within a non-orthogonal multiple access (NOMA) system operating in the finite blocklength (FBL) regime, specifically designed for short-packet communications. By leveraging the effective capacity (EC) framework, latency and reliability in FBL, encompassing parameters such as the block error probability and the delay outage probability, are analyzed for two scenarios, namely, system operation with multiple NOMA pairs and the two-user NOMA operation. Closed-form expressions for the EC in the two cases are derived by assuming that transmissions are subject to Rayleigh fading and adopting a practical path-loss model. Numerical results are provided to validate the analytical findings, and to highlight the impact of the transmit signal-to-noise ratio, the blocklength, the delay exponent, and the block error probability, on the EC and the delay outage probability. Furthermore, various pairing configurations are investigated and demonstrate that the paired NOMA set attains the highest total EC for users experiencing substantial differences in their channel conditions.

**Index Terms**—NOMA, FBL, short-packet communications, effective capacity, delay outage probability, block error probability.

## I. INTRODUCTION

### A. Context and Literature Review

The emergence of 6G wireless communication brings into front a new era of innovative technologies, opening pathways for unparalleled connectivity and extremely high data exchange. In this context, the demand for low-latency communications has increased, prompting the exploration and optimization of communication protocols that can meet the requirements of applications with stringent delay and access requirements [1]–[3]. In particular, non-orthogonal multiple access (NOMA), rate-splitting multiple access (RSMA) and sparse vector coding (SVC), are emerging as key technologies [4]–[6] to enhance the spectral efficiency, achieve better block error rate performance and lower transmission latency, and improve the user fairness, especially in critical use cases of superimposed transmission approach, which enable the simultaneous transmission of multiple signals by superimposing

them in the same frequency band [6]. This approach minimizes transmission delays by avoiding the need for strict scheduling or orthogonal resource allocation, making it particularly suitable for scenarios requiring extreme ultra-reliable low-latency communications (extreme-uRLLC) [7].

In extreme-uRLLC scenarios, where reliability and latency are critical, NOMA presents important advantages over conventional orthogonal multiple access schemes [8]. Firstly, NOMA allows for simultaneous transmissions to multiple users within the same frequency/time resource [9], [10]. This is particularly beneficial in extreme-uRLLC scenarios where spectrum resources are limited, as NOMA enables more efficient utilization of the available bandwidth [4], [11]. Secondly, NOMA offers improved reliability by leveraging power-domain multiplexing, where users with different channel conditions are assigned different power levels. This ensures that the weaker users receive sufficient signal strength to decode their data, thus enhancing their data-reception reliability and reducing the probability of packet loss [8], [12].

On the other hand, SVC is specifically designed for short-packet transmissions and has shown to be very effective in extreme-uRLLC scenarios [13], [14]. While NOMA focuses on efficient resource utilization among multiple users, SVC aims to enhance the reliability and latency performance of individual short-packet transmissions. The SVC technique achieves this by encoding information into sparse vectors [6]. Despite its potential advantages and efficiency, implementing SVC for extreme-uRLLC presents significant challenges. For example, achieving the ultra-low latency required for extreme-uRLLC while maintaining exceptionally high reliability, typically 99.999% or higher, demands efficient encoding and decoding algorithms that can process information rapidly without compromising error correction. The challenge is further compounded by the short packet nature of extreme-uRLLC transmissions, which limits the amount of data available for processing and error protection. Therefore, resource allocation becomes complex since SVC must coexist with other services, while efficiently utilizing the radio resources. Another important challenge in SVC implementation is the energy efficiency, which is crucial especially for IoT devices. This necessitates careful consideration of the power consumption and maintaining low computational complexity for real-time processing on the resource-constrained devices. Overcoming these challenges requires innovative approaches in coding scheme design, resource allocation algorithms, and network architectures, to fully harness the SVC's potential for extreme-uRLLC and short-packet applications.

Taking into account the above discussions, by efficiently allocating the available resources and adapting to the varying

Manuscript received April 22, 2024; revised September 16, 2024, January 01, 2025, and April 09, 2025; accepted April 19, 2025. The associate editor coordinating the review of this article and approving it for publication was Dr. Swades De.

Z. Mohamed and S. Aïssa are with the Institut National de la Recherche Scientifique (INRS), University of Quebec, Montreal, QC, H5A 1K6, Canada (email: {zina.mohamed, sonia.aïssa}@inrs.ca.). M. Amjad is with the School of Engineering, Computing and Mathematical Sciences, University of Wolverhampton, WV1 1LY, UK. (email: m.amjad6@wlv.ac.uk). L. Musavian is with the School of Computer Science and Electronic Engineering, University of Essex, CO4 3SQ, UK (email: leila.musavian@essex.ac.uk).

Part of this work appears in the 2020 IEEE International Symposium on Personal, Indoor and Mobile Radio Communications.

quality of service (QoS) requirements, NOMA can effectively meet the stringent reliability and latency demands of extreme-uRLLC services, making it a promising candidate for next-generation wireless communication networks [15].

The working principle of NOMA with short-packet communication is similar to that of conventional NOMA, using the superposition coding (SC) at the transmitter and successive interference cancellation (SIC) at the receiver [16]–[19]. In practice, the SIC consists in decoding the multi-user signals with the strongest received signal power first, subtracting them from the combined signal, and then decoding the difference as the weakest signal power. The process is repeated until decoding the desired signal is complete. While imperfect SIC is a critical challenge in NOMA systems [20], particularly in the context of finite blocklength (FBL) and short-packet communication, its impact can vary based on system conditions. Imperfect SIC introduces residual interference that can degrade performance, increase interference [21], and reduce reliability [22], which is especially concerning in short-packet communications where the tolerance to errors is limited. However, under specific conditions, the impact of imperfect SIC can be mitigated or even become negligible. For instance, in low signal-to-noise ratio (SNR) environments, noise dominates over interference, reducing the influence of SIC imperfections. Additionally, the short-packet lengths limit the time for interference accumulation, minimizing the effect of residual interference [23]. Furthermore, advanced receiver techniques, optimal power allocation strategies, and user pairing schemes are effective in mitigating the impact of imperfect SIC. These approaches allow well-designed NOMA systems with short-packet communication to maintain their performance advantages, ensuring reliability and efficiency in next-generation wireless networks.

Generally speaking, the challenge in short-packet communication lies in optimizing the system performance while adhering to the stringent latency constraints. From a theoretical perspective, the foundation of conventional wireless communication is Shannon capacity [24], which is an asymptotic metric that represents the maximum data rate that can be achieved while ensuring a specific level of reliability. In short-packet communication, this metric falls short in capturing the complex interplay between reliability, latency, and throughput [24], [25]. Indeed, in the pioneering work of [24], which investigated the achievable rate of a FBL communication link constrained by a given error probability under additive white Gaussian noise (AWGN), it was proven that Shannon's formula cannot be used to approximate the maximum achievable rate with a FBL. This work defined a penalty factor related to channel capacity and error probability, and introduced them into the process of calculating the achievable rates in FBL regime. Extending the study to the case of Rayleigh block fading channels, the work in [26] examined the trade-off between reliability, latency, and throughput in FBL regime, and emphasized the importance of short-packet communication to meet the low-latency requirements.

To adhere to the delay outage probability constraint and assess the suitability of short-packet communication for achieving low delay, it is crucial to adopt a straightforward and adaptive approach. In this regard, Shannon capacity and ef-

fective capacity (EC) are two important concepts that serve different purposes in understanding and designing wireless communication systems. Shannon capacity, derived from the Shannon-Hartley theorem, represents the theoretical maximum rate at which information can be reliably transmitted over a communication channel with arbitrarily small error probability. It is calculated using bandwidth and signal-to-noise ratio, assuming ideal conditions. Unlike Shannon capacity, the EC considers QoS requirements, such as the delay constraint, making it more suitable for time-varying channels commonly encountered in wireless communications [8], [27], [28]. It provides statistical guarantees on QoS metrics. Additionally, the EC is particularly useful for practical system design and performance analysis, especially providing less complicated closed-form formulae for the performance metrics in systems with statistical delay constraints, as compared to those obtained by the Shannon-Hartley theorem. Besides, the EC framework helps in optimizing the resource allocation and transmission strategies while meeting specific QoS requirements, which is crucial for applications with statistically strict latency constraints or limited energy resources. In the context of NOMA systems operating in the FBL regime, the EC analysis is crucial since it provides a comprehensive framework to evaluate the maximum sustainable constant arrival rate while satisfying statistical QoS requirements, making it especially relevant for delay-sensitive applications in extreme-uRLLC scenarios [29]–[31]. Unlike traditional capacity analysis, EC captures the impact of physical-layer parameters (such as channel conditions and transmission power) and link-layer QoS constraints (such as delay and buffer overflow probabilities). By conducting EC analysis for NOMA with FBL, this paper aims to provide valuable insights into the design and performance evaluation of next-generation wireless networks that must support diverse services with varying latency requirements [32]. By characterizing the maximum arrival rate while ensuring compliance with specified delay constraints, EC provides a holistic perspective on system performance. Precisely, this framework represents a dual concept of effective bandwidth, offering insights into how to achieve optimal throughput while maintaining low latency. Thus, the adoption of the EC framework in NOMA and FBL communication promises to unlock new avenues for designing efficient wireless systems tailored to the demands of modern applications with extreme-uRLLC requirements [8].

So far, several studies have been conducted on the combination of NOMA and FBL communication as a mean to realize low latency [8], [27], [33], [34]. Specifically, [8] investigated the performance of NOMA with FBL when considering the amount of physical-layer transmission latency, and showed that the observed latency under reliability constraint is lower compared to the system counterpart with OMA. Challenges associated with the SIC and transmission rate while using FBL were highlighted in [33]. In [27], the EC of NOMA with FBL communication was investigated, taking into consideration the delay exponent and block error probability of the system. Subsequently, novel closed-form and asymptotic formulae for the EC of arbitrarily ordered NOMA users with FBL were derived. Besides, the work in [34] investigated the max-min fairness in uplink NOMA with FBL, specifically looking at a

joint power control and transmission rate optimization aiming to maximize the users' minimum effective throughput under reliability constraints, where effective throughput is taken as a balanced metric that incorporates both the transmission rate and the error rate.

## B. Contributions

Motivated by the principles of short-packet communication and the potential of NOMA, this paper introduces a framework aimed at investigating and assessing the performance of NOMA with FBL in terms of the achievable EC. Assuming  $V$  users in the system, two NOMA scenarios are considered: the first involves two paired users, and the second involves multiple NOMA pairs. The detailed contributions of this research work can be summarized as follows: **(i)** Providing the closed-form expression for the probability density function (PDF) of independent but not necessarily identically distributed (i.n.i.d.) ordered exponential random variables (RVs). These RVs represent the received SNRs at the users. **(ii)** Using the obtained PDF, closed-form expressions for the achievable EC in the scenarios considered in this work, i.e., the two-user NOMA and NOMA with multiple pairs, are derived considering FBL transmissions. To the best of the authors' knowledge, no previous work has established these closed-form expressions when the channels are i.n.i.d., which is the case in this study. Given the complexity of the derived formulae, we also present simplified closed-form expressions for the EC at high transmit SNRs and at extremely high SNRs. **(iii)** Based on the developed analysis, Monte-Carlo simulations are conducted to confirm the correctness of the derivations. Moreover, the impact of the packet blocklength, transmit SNR, delay exponent, and block error probability, on the achievable EC and the delay outage probability, are investigated. The findings confirm the correctness of the derivations, and show that the achievable EC increases with the transmit SNR, and that the delay outage probability decreases as the delay exponent increases in the two-user NOMA. It is also shown that the total EC of a NOMA pair with users having highly distinct channel conditions is higher compared to when users have less distinct channel conditions.

The rest of the paper is organized as follows. First, the short-packet communication model is detailed in section II. In section III, the concepts related to the theory of EC and the proposed expressions for the EC of NOMA with FBL are presented. Numerical results with their insights are discussed in section IV. Finally, the paper is concluded in section V.

*Notations:*  $Q^{-1}(\cdot)$  is the inverse of the Gaussian Q-function  $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{w^2}{2}} dw$ ,  $\mathbb{E}[\cdot]$  denotes mathematical expectation,  $\Gamma(\cdot)$  is the Gamma function,  $\Gamma(\cdot, \cdot)$  is the incomplete Gamma function,  $\Pr\{a>b\}$  is the probability that  $a>b$  holds.  $H(a, b, z)$  is the confluent hypergeometric function of the second kind defined by  $H(a, b, z) = \frac{1}{\Gamma(a)} \int_0^\infty e^{-zt} t^{a-1} (1+t)^{b-a-1} dt$ , for  $\text{Re}[a], \text{Re}[z] > 0$ , [35, Eq. (16)],  $\text{Ei}(x) = -\int_{-x}^\infty \frac{e^{-w}}{w} dw$  is the exponential integral [36, Eq. (12.78)],  $G_{\dots}^{\dots}(\cdot | \cdot | \cdot | x, z)$  is the bivariate Meijer-G function, and  ${}_1F_1(a, b, z)$  is the confluent hyper-

geometric function of the first kind. Table 1 summarizes the main notations that are used in the paper.

Table I  
SUMMARY OF THE MAIN SYMBOLS AND NOTATIONS

$V$	total number of users	$\epsilon$	block error probability
$\alpha_i$	power allocation coefficient of user $i$	$r_i$	service rate for user $i$
$s_i(\tau)$	message intended for user $i$ at time $\tau$	$\delta_i$	channel dispersion of user $i$
$P$	total transmit power at the BS	$T_{ec}$	total effective capacity
$n$	blocklength	$a_i(\tau)$	number of queue packets at time $\tau$
$L_i(\tau)$	channel coefficient between the BS and user $i$ at time $\tau$	$\theta_i$	delay exponent of user $i$
$y_k$	receive signal at user $k \in \{k_1, k_2\}$	$q_i(\infty)$	steady-state of transmit buffer
$n_k$	AWGN	$D_{\max}^i$	maximum delay
$k_1$	strong user	$C_k$	effective capacity for user $k \in \{k_1, k_2\}$
$k_2$	weak user	$N_oB$	noise power
$\text{SNR}_{k_1}$	received SNR at the strong user	$\text{SINR}_{k_2}$	received SINR at the weak user
$\phi$	combination of all NOMA pairs	$f_{(i:V)}$	PDF of ordered channel gains

## II. SYSTEM MODEL

The system under study is a power-domain downlink NOMA system with short-packet communications between the base station (BS) and  $V$  single-antenna users, as shown in Fig. 1. Let  $L_i(\tau)$  denote the channel coefficient between the BS and user  $i \in \{1, \dots, V\}$  at time  $\tau$ .<sup>1</sup> The channel coefficient of the link between the BS and a user  $i \in \{1, \dots, V\}$  includes the effects of the large- and small-scale fading. The quasi-static fading model, reflected in  $h_i(\tau)$ , is used to model the small-scale fading, which remains invariant during a packet transmission under the assumption that the length of the packet is sufficiently short [37], [38], and follows a Rayleigh distribution with zero mean and unit variance. The large-scale fading follows a practical path-loss model,  $d_i^{-\frac{L_P}{2}}$ , where  $d_i$  is the distance between the BS and the user, and  $L_P$  is the path-loss exponent. As such, the channel coefficient of the link between the BS and user  $i$  is given by  $L_i(\tau) = h_i(\tau) / d_i^{\frac{L_P}{2}}$ .

The basic operation of the communication model is summarized as follows: the upper-layer packets of each user are assembled into frames, then stored in the transmission buffer of the BS, and later transmitted over the wireless channel as bit streams (cf. Fig. 1). Without loss of generality, we assume that the NOMA users are classified based on their channel conditions and that their corresponding instantaneous channel state information (CSI),  $L_i(\tau) \forall i \in \{1, \dots, V\}$ , is available at the BS. Hence, we have  $|L_1(\tau)|^2 \leq |L_2(\tau)|^2 \leq \dots \leq |L_V(\tau)|^2$ . The BS recognizes the user ordering based on the CSI, and allocates power to different user streams from a set of fixed power factors. Assume that the set of  $V$  users is partitioned into  $\frac{V}{2}$  pairs, and denote by  $\phi = \{\phi_1, \phi_2, \dots, \phi_{\frac{V}{2}}\}$  the combination of all NOMA

<sup>1</sup>For simplicity, the time index  $\tau$  will hereafter be omitted whenever it is clear from the context.

pairs. Considering the  $m^{\text{th}}$  NOMA pair, the strong user is denoted by  $k_1$  and the weak user is denoted by  $k_2$ , such that  $\phi_m = \{(k_2, k_1) \mid k_2 \neq k_1, |L_{k_2}|^2 \leq |L_{k_1}|^2, \forall m \in \mathbb{M}\}$ , in which  $\mathbb{M} = \{1, 2, \dots, \frac{V}{2}\}$  is the set of indices of the  $\frac{V}{2}$  pairs.

#### A. Received Signal

Considering that transmissions are of FBL type, we investigate the achievable EC of users  $k_1$  and  $k_2$  in the  $m^{\text{th}}$  NOMA pair. It is noteworthy that the inter-pair multiple access is based on frequency-division multiple access (FDMA).

The received signals at the paired users  $k_1$  and  $k_2$  of the  $m^{\text{th}}$  NOMA pair can then be formulated as

$$y_k = \sum_{u=1}^U L_k \sqrt{\alpha_u P} s_u + n_k, \quad \forall k \in \{k_1, k_2\}, \quad (1)$$

where the  $\alpha$ 's are the power allocation coefficients satisfying  $\alpha_{k_1} + \alpha_{k_2} = 1$ ,  $P$  is the total transmit power of the BS,  $s_k$  is the message intended for user  $k$ , and  $n_k$  represents its corresponding AWGN.  $U$  is the number of users per NOMA pair.

The strong user,  $k_1$ , performs SIC and detects the message of the weak user,  $k_2$ . Specifically, the so-called strong user will remove the weak-user's message from its received message. Assuming perfect SIC, the received SNR of user  $k_1$  will be given by

$$\begin{aligned} \text{SNR}_{k_1} &= \alpha_{k_1} \rho |L_{k_1}|^2 \\ &= \alpha_{k_1} \gamma_{k_1}, \end{aligned} \quad (2)$$

where  $\rho$  is the transmit SNR, i.e.,  $\rho = \frac{P}{N_0 B}$ , in which  $N_0 B$  denotes the noise power and  $B$  is the bandwidth.

At the weak-user's side, i.e., node  $k_2$ , the message of user  $k_1$  is treated as noise. Therefore, the weak user will only decode its own message. Hence, the signal-to-interference-plus-noise ratio (SINR) of the weak user is expressed as

$$\begin{aligned} \text{SINR}_{k_2} &= \frac{\alpha_{k_2} \rho |L_{k_2}|^2}{\alpha_{k_1} \rho |L_{k_2}|^2 + 1} \\ &= \frac{\alpha_{k_2} \gamma_{k_2}}{\alpha_{k_1} \gamma_{k_2} + 1}. \end{aligned} \quad (3)$$

Let  $\epsilon$  be the block error probability. As we consider the communications with short packets, the Shannon's formula cannot be used to express the users' achievable rates [25]. With FBL, the achievable rates of users  $k_1$  and  $k_2$  can be expressed as a function of not only the received SNR (or SINR), but also the block error probability  $\epsilon$  and the transmission blocklength  $n$ . Based on [38], these rates can be approximated, in the unit bit/s/Hz, as follows:

$$r_{k_1} = \frac{2}{V} \left( \ln(1 + \alpha_{k_1} \gamma_{k_1}) - \sqrt{\frac{\delta_{k_1}}{n}} Q^{-1}(\epsilon) \right), \quad (4)$$

$$r_{k_2} = \frac{2}{V} \left( \ln \left( 1 + \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right) - \sqrt{\frac{\delta_{k_2}}{n}} Q^{-1}(\epsilon) \right), \quad (5)$$

where  $\delta_{k_1} = 1 - (1 + \alpha_{k_1} \gamma_{k_1})^{-2}$  and  $\delta_{k_2} = 1 - \left(1 + \frac{\alpha_{k_2} \gamma_{k_2}}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{-2}$  are the channel dispersions pertaining to users  $k_1$  and  $k_2$ , respectively,  $\gamma_{k_1} = \rho |L_{k_1}|^2$  and  $\gamma_{k_2} = \rho |L_{k_2}|^2$ .

### III. EFFECTIVE CAPACITY FRAMEWORK

#### A. Preliminaries on the Theory of Effective Capacity

Proposed in [39], the EC is defined as the dual of the effective bandwidth, and introduces a link-layer QoS metric, such as the delay outage probability and the probability of non-empty buffer. Hereafter, main concepts of the EC theory are presented, as they are needed later for the core of the paper's contribution.

Let  $\theta_i$  denote the delay exponent of user  $i \in \{1, \dots, V\}$ ,  $q_i(\infty)$  be the steady state of  $i^{\text{th}}$  buffer [40], [41], and the link capacity (service process) at time  $\tau$  be  $r_i(\tau)$ . In practice, buffer overflow occurs when  $q_i(\infty)$  exceeds the maximum length of the buffer. Assume  $x$  to be a maximum threshold on  $q_i(\infty)$ . Then, using the large deviation theorem [42], we can write

$$-\lim_{x \rightarrow \infty} \frac{\ln(\Pr\{q_i(\infty) > x\})}{x} = \theta_i. \quad (6)$$

For a target  $\theta_i$ , the buffer overflow probability, given in (6), can be satisfied if

$$\Lambda_{a_i}(\theta_i) + \Lambda_{r_i}(-\theta_i) = 0, \quad (7)$$

where  $\Lambda_{a_i}(\theta_i) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln(\mathbb{E}[e^{\theta_i \sum_{\tau=1}^T a_i(\tau)}])$  is the Gärtner-Ellis limit of the source process (arrival rate),  $\Lambda_{r_i}(\theta_i) = \lim_{T \rightarrow \infty} \frac{1}{T} \ln(\mathbb{E}[e^{\theta_i \sum_{\tau=1}^T r_i(\tau)}])$  is the Gärtner-Ellis limit of the service process [41], [43],  $a_i(\tau)$  is the number of arriving packets and  $r_i(\tau)$  is the link capacity, at time  $\tau$ . Suppose that the source rate  $a_i(\tau)$  is constant, such that  $a_i(\tau) = a_i$ . From (7), we can get the maximum arrival rate for some unique  $\theta_i$  (delay QoS exponent). This rate is named EC and can be approximated by  $-\frac{\Lambda_{r_i}(-\theta_i)}{\theta_i}$  [39].

From (6), the delay experienced by the source packets in the  $i^{\text{th}}$  buffer at time  $\tau$  can be estimated in terms of the delay outage probability [39]. That is,

$$\Pr\{D_i(\tau) > D_{\max}^i\} \approx \Pr\{q_i(\infty) > 0\} e^{-\theta_i \mu_i D_{\max}^i}, \quad (8)$$

where  $\Pr\{q_i(\infty) > 0\}$  represents the probability of non-empty buffer,  $D_{\max}^i$  is the maximum delay, and  $\mu_i = C_i$  is the EC satisfying a certain QoS metric for user  $i$  [39]. As per (8), the value of  $\theta_i$ , recalling that  $\theta_i > 0$ , is the decay rate of the outage probability corresponding to user  $i$ . A more stringent delay requirement can be represented with a larger value of  $\theta_i$ , while a smaller value of  $\theta_i$  shows a less stringent delay requirement.

#### B. Effective Capacity with FBL

Based on the EC framework, our goal is to investigate the latency performance of the NOMA system with short-packet transmissions. As mentioned earlier, the traditional stochastic model for deriving the achievable EC based on the Shannon limit as the service rate cannot be adopted when considering FBL transmissions. Therefore, for the service rates of our NOMA system with short-packet communication, we use  $r_{k_1}$  and  $r_{k_2}$  as provided in (4) and (5). Based on the stochastic model for the achievable EC with short-packet communication provided in [44] and following the service rates derived in (4)

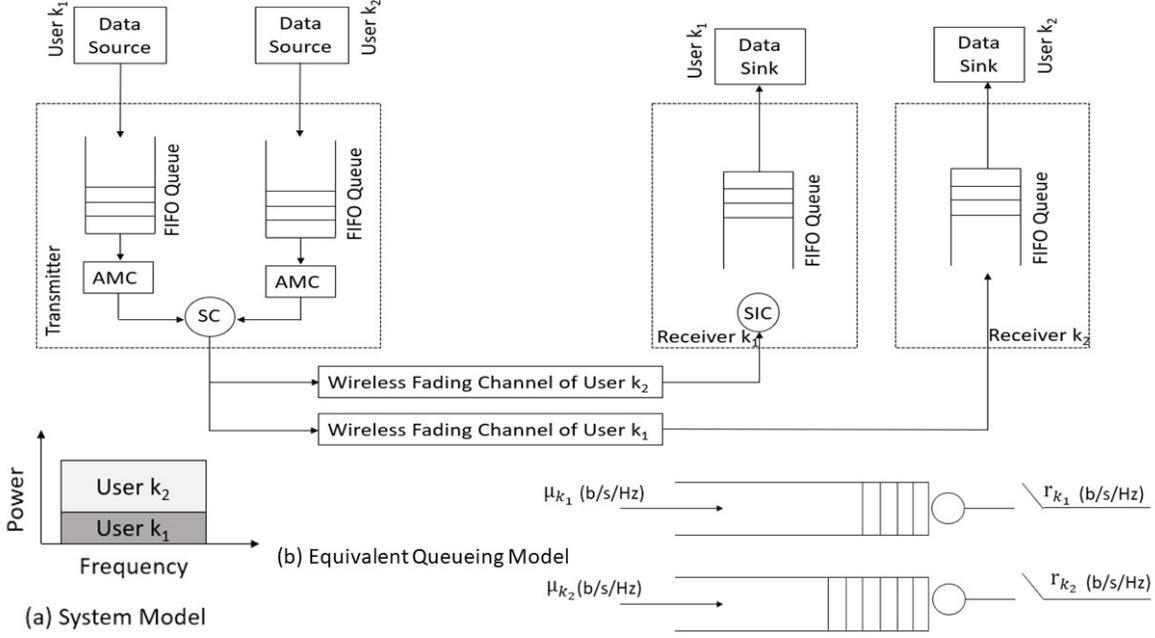


Figure 1. Two-user NOMA operation with FBL with their respective queues: (a) describes the system model with two queues at the BS with their respective receivers, and (b) depicts the equivalent queueing model with the arrival rate and service rate.

and (5), the EC of user  $k \in \{k_1, k_2\}$  in the  $m^{\text{th}}$  NOMA pair with FBL, denoted  $C_k$ , is formulated as

$$C_k = -\frac{1}{\theta_k n} \ln \left( \mathbb{E} \left[ \epsilon_k + (1 - \epsilon_k) e^{-\theta_k n r_k} \right] \right), \quad (9)$$

where  $\theta_k$  is the QoS constraint of user  $k$ , and  $\epsilon_k$  is the block error probability of user  $k$ .

To develop further, the achievable EC can be defined as

$$C_k = -\frac{1}{\theta_k n} \times \ln \left( \int_0^\infty (\epsilon_k + (1 - \epsilon_k) e^{-\theta_k n r_k}) f_{k:V}(\gamma_k) d\gamma_k \right), \quad (10)$$

where  $f_{k:V}(\gamma_k)$  is the PDF of the  $k^{\text{th}}$  order statistic of the SNR/SINR at user  $k \in \{k_1, k_2\}$ . To derive the closed-form expression of the EC, the said PDF is required. In the following, a closed-form expression for the PDF of i.n.i.d. exponential RVs is presented.

### C. Fundamental Statistics: PDF of Ordered Statistics

Let the received SNRs at the users  $i \in \{1, \dots, V\}$ , denoted  $\gamma_i = \rho |h_i|^2 / d_i^{L_P}$  for  $i \in \{1, \dots, V\}$ , be i.n.i.d. RVs having cumulative distribution functions (CDFs) denoted by  $F_1(\gamma), F_2(\gamma), \dots, F_V(\gamma)$ , and PDFs denoted by  $f_1(\gamma), f_2(\gamma), \dots, f_V(\gamma)$ . Let  $\gamma_{1:V} \leq \gamma_{2:V} \leq \dots \leq \gamma_{V:V}$  denote the order statistics obtained by arranging the  $\gamma_i$ 's in increasing order. Then, the PDF of the  $i^{\text{th}}$  order statistic  $\gamma_{i:V}$ , can be written as [45]

$$f_{i:V}(\gamma) = \frac{1}{(i-1)!(V-i)!} \sum_p F_{j_1}(\gamma) \cdots F_{j_{i-1}}(\gamma) f_{j_i}(\gamma) (1 - F_{j_{i+1}}(\gamma)) \cdots (1 - F_{j_V}(\gamma)), \quad (11)$$

where  $\sum_p$  denotes the summation over all  $V!$  permutations  $(j_1, j_2, \dots, j_V)$  of  $\{1, 2, \dots, V\}$ . Using matrix permanent, the PDF can be recast as follows [45]:

$$f_{i:V}(\gamma) = \frac{1}{(i-1)!(V-i)!} \text{per} \begin{bmatrix} F_1(\gamma) & f_1(\gamma) & 1 - F_1(\gamma) \\ \vdots & \vdots & \vdots \\ F_V(\gamma) & f_V(\gamma) & 1 - F_V(\gamma) \end{bmatrix}, \quad (12)$$

and the CDF can be expressed as

$$F_{i:V}(\gamma) = \sum_{j=i}^V \frac{1}{j!(V-j)!} \text{per} \begin{bmatrix} F_1(\gamma) & 1 - F_1(\gamma) \\ \vdots & \vdots \\ F_V(\gamma) & 1 - F_V(\gamma) \end{bmatrix}. \quad (13)$$

In this paper, we consider the case when the variables  $\gamma_i$ , for  $i \in \{1, \dots, V\}$ , are i.n.i.d. exponential RVs, with PDF given by  $f_i(\gamma) = \frac{1}{\rho_i} e^{-\frac{\gamma}{\rho_i}}$ , where  $\rho_i = \rho / d_i^{L_P}$ , and CDF expressed as  $F_i(\gamma) = 1 - e^{-\frac{\gamma}{\rho_i}}$ . Therefore,  $f_i(\gamma) = \frac{1}{\rho_i} (1 - F_i(\gamma)) = \frac{1}{\rho_i} e^{-\frac{\gamma}{\rho_i}}$ ,  $\forall i \in \{1, \dots, V\}$ , with  $\rho_i \geq 0$  and  $\gamma > 0$ .

Using Eq. (12) and based on [46], the PDF of the  $i^{\text{th}}$  order statistic  $\gamma_{i:V}$  can then be expressed as

$$f_{i:V}(\gamma) = \frac{1}{(i-1)!(V-i)!} \times \sum_p \prod_{a=1}^{i-1} \left( 1 - e^{-\frac{\gamma}{\rho_a}} \right) \frac{1}{\rho_i} e^{-\frac{\gamma}{\rho_i}} \prod_{c=i+1}^V e^{-\frac{\gamma}{\rho_c}}, \quad (14)$$

where  $\sum_p = \sum_{\text{all permutations } V!}$ .

Next, we provide closed-form expressions for the achievable EC in the NOMA system operating in FBL regime. First, we focus on the two-user NOMA scenario.

### D. Achievable Effective Capacity with FBL: Two-User NOMA

In this case, we assume that two users among the total of  $V$  users are paired, namely, the strong user and the weak user, denoted by  $k_1$  and  $k_2$ , respectively. Therefore, the transmission rates of these paired users can be recast as

$$r_{k_1} = \ln(1 + \alpha_{k_1} \gamma_{k_1}) - \sqrt{\frac{\delta_{k_1}}{n}} Q^{-1}(\epsilon_{k_1}), \quad (15)$$

$$r_{k_2} = \ln\left(1 + \frac{\alpha_{k_2} \gamma_{k_2}}{\alpha_{k_1} \gamma_{k_2} + 1}\right) - \sqrt{\frac{\delta_{k_2}}{n}} Q^{-1}(\epsilon_{k_2}), \quad (16)$$

where  $\gamma_k = \rho |h_k|^2 / d_k^{L_P}$  and  $\alpha_k$  is the power coefficient, for  $k \in \{k_1, k_2\}$ .

1) *EC of the NOMA Strong User:* The closed-form expression for the achievable EC of the NOMA strong user ( $k_1$ ) in FBL can be approximated as shown in Eq. (17), in which  $\sum_p = \sum_{\text{all permutations } V!}$ ,  $\nu_1 = \sum_{c=k_1+1}^V \left(\frac{1}{\rho_c}\right) + \frac{1}{\rho_{k_1}}$ ,  $a_{11} = 2\zeta_{k_1} - k_1 + 1$ , and  $a_{12} = 2\zeta_{k_1} - k_1 - 1$ .

Appendix A provides more details on the derivation of the closed-form expression for the achievable EC of the strong user.

2) *EC of the NOMA Weak User:* For the NOMA weak user ( $k_2$ ), the closed-form expression of the EC in FBL is shown in Eq. (18), in which  $\sum_p = \sum_{\text{all permutations } V!}$  and  $\nu_2 = \sum_{c=k_2+1}^V \frac{1}{\rho_c} + \frac{1}{\rho_{k_2}}$ .

Further details regarding the derivation process of the closed-form expression for the EC of the weak user are provided in Appendix B.

3) *EC at High Transmit SNRs:* In this part, we investigate the performance of the two-user NOMA in FBL regime by leveraging the closed-form expressions for the EC of the strong user and weak user at high SNR values. Indeed, the derived closed-form expressions reported in (17) and (18) are quite complex, which can make it difficult to gain direct analytical insights from these expressions. For further insights, various approximations are leveraged to simplify the achievable EC formulation and their corresponding closed-form expressions for the two-user NOMA in FBL. For this purpose, the channel dispersion  $\delta_i$ ,  $i \in \{1, \dots, V\}$ , is approximated as  $\delta_i \approx 1$  for high SNR values [25]. Finally, using this approximation, the EC attained by the strong user at high transmit SNR simplifies to

$$\bar{C}_{k_1} = -\frac{\ln\left(\mathbb{E}\left[\epsilon_{k_1} + (1 - \epsilon_{k_1})(1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1}}\right]\right)}{\theta_{k_1} n}, \quad (19)$$

and the weak-user's achievable EC at high transmit SNR becomes

$$\bar{C}_{k_2} = -\frac{\ln\left(\mathbb{E}\left[\epsilon_{k_2} + (1 - \epsilon_{k_2})\left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{2\zeta_{k_2}} e^{\beta_{k_2}}\right]\right)}{\theta_{k_2} n}. \quad (20)$$

These equations can be further simplified by deriving their closed-form expressions. To achieve this, the order statistics from (11) are used first. Consequently, the achievable ECs of

the strong and weak users at high transmit SNR values can respectively be expanded as

$$\bar{C}_{k_1} = -\frac{1}{\theta_{k_1} n} \ln\left(\int_0^\infty \left(\epsilon_{k_1} + (1 - \epsilon_{k_1})(1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1}}\right) \times f_{k_1:V}(\gamma_{k_1}) d\gamma_{k_1}\right). \quad (21)$$

$$\bar{C}_{k_2} = -\frac{1}{\theta_{k_2} n} \ln\left(\int_0^\infty \left(\epsilon_{k_2} + (1 - \epsilon_{k_2})\left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{2\zeta_{k_2}} \times e^{\beta_{k_2}}\right) f_{k_2:V}(\gamma_{k_2}) d\gamma_{k_2}\right). \quad (22)$$

The integrals in Eqs. (21) and (22) can be derived by repeating the processes outlined in Appendix A and Appendix B. Finally, the closed-form expressions for the achievable ECs of the strong and weak users at high SNRs are obtained as shown in Eq. (23) and Eq. (24), in which  $\sum_p = \sum_{\text{all permutations } V!}$ .

Compared to the closed forms shown in (17) and (18), the closed-form expressions obtained in Eq. (23) and (24) are simpler and more suitable for investigating the performance of the two-user NOMA with short-packet communication.

4) *EC at Extremely High Transmit SNRs:* Here, the impact of extremely high transmit SNR on the achievable EC of two-user NOMA in FBL is studied. For this purpose, the achievable EC of the NOMA strong and weak users at extremely high transmit SNR can be derived by considering  $\rho \rightarrow \infty$  in the formulations. Accordingly, the EC of the strong and weak users at extremely high SNR are respectively given as

$$\lim_{\rho \rightarrow \infty} -\frac{1}{\theta_{k_1} n} \ln\left(\mathbb{E}\left[\epsilon_{k_1} + (1 - \epsilon_{k_1}) \times e^{-\theta_{k_1} n \ln(1 + \alpha_{k_1} \gamma_{k_1}) - \sqrt{\frac{1 - (1 + \alpha_{k_1} \gamma_{k_1})^{-2}}{n}} Q^{-1}(\epsilon_{k_1})}\right]\right) = -\frac{1}{\theta_{k_1} n} \ln(\epsilon_{k_1}). \quad (25)$$

$$\lim_{\rho \rightarrow \infty} -\frac{1}{\theta_{k_2} n} \ln\left(\mathbb{E}\left[\epsilon_{k_2} + (1 - \epsilon_{k_2})\left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{2\zeta_{k_2}} \times e^{\beta_{k_2} \sqrt{1 - \left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{-2}}}\right]\right) = -\frac{1}{\theta_{k_2} n} \ln\left(\epsilon_{k_2} + (1 - \epsilon_{k_2}) \alpha_{k_1}^{-2\zeta_{k_2}} e^{\beta_{k_2} \sqrt{1 - \alpha_{k_1}^2}}\right). \quad (26)$$

According to the above, we note that the EC under extremely high transmit SNR is upper bounded by a fixed value, which is not a function of  $\rho$ . In particular, the achievable EC of the strong user is limited by the delay exponent, the block error probability, and the blocklength. Moreover, the achievable EC of the weak user is limited by the quantity  $-\frac{1}{\theta_{k_2} n} \ln\left(\epsilon_{k_2} + (1 - \epsilon_{k_2}) \alpha_{k_1}^{-2\zeta_{k_2}} e^{\beta_{k_2} \sqrt{1 - \alpha_{k_1}^2}}\right)$  as can be obtained from (26) when  $\rho$  is extremely high. Specifically, the weak-user's achievable EC is upper bounded by the transmission penalty, a well-established concept that captures the

$$C_{k_1} = -\frac{1}{\theta_{k_1} n} \ln \left( \frac{\sum_p \frac{\Gamma(k_1)}{\prod_{a=1}^{k_1-1} \rho_a}}{(k_1-1)!(V-u)!\rho_{k_1}} \left( \frac{\epsilon_{k_1}}{\nu_1^{k_1+1}} + (1-\epsilon_{k_1}) \left( \alpha_{k_1}^{k_1} \left( \mathbb{H}(k_1, a_{11}, \alpha_{k_1} \nu_1) + \left( \beta_{k_1} + \frac{\beta_{k_1}^2}{4} \right) \mathbb{H}(k_1, a_{11}, \alpha_{k_1} \nu_1) \right. \right. \right. \right. \right. \\ \left. \left. \left. \left. + \left( \frac{\beta_{k_1}}{2} + \frac{\beta_{k_1}^2}{4} \right) \mathbb{H}(k_1, a_{12}, \alpha_{k_1} \nu_1) \right) \right) \right) \right). \quad (17)$$

$$C_{k_2} = -\frac{1}{\theta_{k_2} n} \ln \left( \frac{\sum_p}{(k_2-1)!(V-k_2)!\rho_{k_2}} \left( \epsilon_{k_2} + (1-\epsilon_{k_2}) \left( \alpha_{k_1}^{-2\zeta_{k_2}} \right. \right. \right. \\ \times \left( \left( \frac{\Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a \nu_2^{k_2+1}} + \frac{2\zeta_{k_2} (\alpha_{k_1} - 1) \alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma\left(1-k_2, \frac{\nu_2}{\alpha_{k_1}}\right) e^{\frac{\nu_2}{\alpha_{k_1}}}}{\prod_{a=1}^{k_2-1} \rho_a} + \frac{\sum_{s=2}^{\infty} \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s}{\prod_{a=1}^{k_2-1} \rho_a} \right. \right. \\ \times \left. \left. \left( \frac{\alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma(s-k_2) {}_1F_1\left(k_2, k_2-s+1, \frac{\nu_2}{\alpha_{k_1}}\right)}{\Gamma(s)} + \alpha_u^{-s} \nu_2^{s-k_2} \Gamma(k_2-s) {}_1F_1\left(s, 1-k_2+s, \frac{\nu_2}{\alpha_{k_1}}\right) \right) \right) \right) \\ + \left( \beta_{k_2} + \frac{\beta_{k_2}^2}{2} \right) \left( \frac{\Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a \nu_2^{k_2+1}} + \frac{2\zeta_{k_2} \prod_{a=1}^{k_2-1} \rho_a}{(\alpha_{k_1} - 1) \alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma\left(1-k_2, \frac{\nu_2}{\alpha_{k_1}}\right) e^{\frac{\nu_2}{\alpha_{k_1}}}} + \frac{\sum_{s=2}^{\infty} \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}}}{\prod_{a=1}^{k_2-1} \rho_a} \right. \\ \times \left. \left( \frac{\alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma(s-k_2) {}_1F_1\left(k_2, k_2-s+1, \frac{\nu_2}{\alpha_{k_1}}\right)}{\Gamma(s)} + \alpha_{k_1}^{-s} \nu_2^{s-k_2} \Gamma(k_2-s) {}_1F_1\left(s, 1-k_2+s, \frac{\nu_2}{\alpha_{k_1}}\right) \right) \right) \\ + \left( \frac{\beta_{k_2}}{2} - \frac{\beta_{k_2}^2}{2} \right) \left( \Gamma(k_2) \nu_2^{-k_2} \left( 1-k_2 (\alpha_{k_1} - 1)^2 + (\alpha_{k_1} - 1) k_2 e^{\nu_2} \left( 1-\nu_2-k_2+\alpha_{k_1} (1+\nu_2+k_2) \right) \right) \right. \\ \left. + \frac{2\zeta_{k_2} (\alpha_{k_1} - 1) \Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a} \left( \nu_2^{1-k_2} \left( 1-\alpha_{k_1} + e^{\nu_2} (1+(\alpha_{k_1} - 1)(\nu_2+k_2)) \right) \text{Ei}(k_2, \nu_2) \right) \right) \\ + \frac{\sum_{s=2}^{\infty} \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left(\frac{\nu_2}{2}\right)^{-k_2}}{\prod_{a=1}^{k_2-1} \rho_a \sqrt{\pi} \Gamma(-2+s)} G_{2,0:1,1:1,1}^{2,0:1,1:1,1} \left( \begin{matrix} 1-k_2, \frac{1}{2}-k_2 \\ 1-k_2 \end{matrix} \middle| \begin{matrix} -1 \\ 0 \end{matrix} \middle| \begin{matrix} 3-s \\ 0 \end{matrix} \middle| \begin{matrix} \frac{2}{\nu_2}, \frac{2\alpha_{k_1}}{\nu_2} \end{matrix} \right) \right) \right). \quad (18)$$

$$\bar{C}_{k_1} \approx -\frac{1}{\theta_{k_1} n} \ln \left( \frac{\sum_p \frac{\Gamma(k_1)}{\prod_{a=1}^{k_1-1} \rho_a}}{(k_1-1)!(V-k_1)!\rho_{k_1}} \left( \frac{\epsilon_{k_1}}{\nu_1^{k_1+1}} + (1-\epsilon_{k_1}) e^{\beta_{k_1}} \mathbb{H}(k_1, a_{11}, \alpha_{k_1} \nu_1) \right) \right). \quad (23)$$

$$\bar{C}_{k_2} \approx -\frac{1}{\theta_{k_2} n} \ln \left( \frac{\sum_p}{(k_2-1)!(V-k_2)!\rho_{k_2}} \left( \frac{\epsilon_{k_2}}{\nu_2^{k_2+1}} + \frac{(1-\epsilon_{k_2}) e^{\beta_{k_2}} \alpha_{k_1}^{-2\zeta_{k_2}} \zeta_{k_2} \sum_{s=2}^{\infty} \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s}{\prod_{a=1}^{k_2-1} \rho_a} \right. \right. \\ \times \left. \left. \left( \frac{\alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma(s-k_2) {}_1F_1\left(k_2, k_1-s+1, \frac{\nu_2}{\alpha_{k_1}}\right)}{\Gamma(s)} + \Gamma(s) \alpha_{k_1}^{-s} (\nu_2)^{s-k_2} \Gamma(k_2-s) {}_1F_1\left(s, 1-k_2+s, \frac{\nu_2}{\alpha_{k_1}}\right) \right) \right) \right). \quad (24)$$

performance degradation experienced by the weak user in a NOMA system due to the short-packet communication, the transmission errors, and the NOMA power allocation in the power-domain NOMA.

#### E. Achievable Effective Capacity with FBL: Multiple NOMA Pairs

Now, we tackle the case of multiple NOMA pairs and investigate the corresponding total achievable EC in the FBL regime. By applying the Gartner-Ellis theorem, and using (4)

and (5) as the transmission rates with FBL, the achievable ECs of the strong and weak users,  $(k_{1_m}, k_{2_m})$ ,  $\forall m \in \mathbb{M} = \{1, 2, \dots, \frac{V}{2}\}$ , in the  $m^{\text{th}}$  NOMA pair can respectively be expressed as:

$$C_{k_{1_m}} = -\frac{1}{\theta_{k_{1_m}} n} \ln \left( \mathbb{E} \left[ \epsilon_{k_{1_m}} + (1-\epsilon_{k_{1_m}}) \right. \right. \\ \left. \left. \times \left( 1 + \alpha_{k_{1_m}} \gamma_{k_{1_m}} \right)^{\frac{4\zeta_{k_{1_m}}}{V m}} e^{\frac{2\beta_{k_{1_m}} \sqrt{\delta_{k_{1_m}}}}{V m}} \right] \right). \quad (27)$$

$$C_{k_{2m}} = \frac{-1}{\theta_{k_{2m}} n} \ln \left( \mathbb{E} \left[ \epsilon_{k_{2m}} + (1 - \epsilon_{k_{2m}}) \left( 1 + \frac{\gamma_{k_{2m}} + 1}{\alpha_{k_{1m}} \gamma_{k_{2m}} + 1} \right)^{\frac{4\zeta_{k_{2m}}}{V_m}} e^{\frac{2\beta_{k_{2m}} \sqrt{\delta_{k_{2m}}}}{V_m}} \right] \right). \quad (28)$$

As observed, the achievable EC of multiple NOMA pairs expressed based on (27) and (28), and the EC of the two-user NOMA shown in (29) and (38), have similar forms. Therefore, based on the steps described in Appendix A and Appendix B, the closed-form expressions for the achievable EC of users  $k_{1m}$  and  $k_{2m}$  in multiple NOMA pairs can be determined.

Finally, the total EC can be calculated as  $T_{ec} = \sum_{m=1}^{\frac{V}{2}} (C_{k_{1m}} + C_{k_{2m}})$ , with  $k_{1m}$  and  $k_{2m}$  denoting the strong and weak users, respectively, in the  $m^{\text{th}}$  NOMA pair, where  $m \in \mathbb{M} = \{1, 2, \dots, \frac{V}{2}\}$ . The analytical results regarding the multiple NOMA pairs will be further investigated in section IV. Specifically, the users with more distinct and less distinct channel conditions will be paired together, and their  $T_{ec}$  will be analyzed with respect to the transmit SNR.

#### IV. NUMERICAL RESULTS AND DISCUSSION

##### A. Simulation Set-up and Parameters

In this section, we evaluate the performance of the NOMA system with finite blocklength and validate the obtained EC formulae. In particular, Monte-Carlo simulations for the two-user NOMA case are performed and the accuracy of the proposed closed-form expressions is confirmed.

Unless otherwise stated, the simulation set-up is as follows:  $V = 10$ , the 2<sup>nd</sup> and 8<sup>th</sup> users are those paired together, such that  $k_2 = 2$  and  $k_1 = 8$ ,  $\alpha_{k_2} = 0.8$ ,  $\alpha_{k_1} = 0.2$ ,  $\theta = 0.01$ ,  $\epsilon = 10^{-5}$  and  $n = 100$ . We assume that  $\epsilon$  is the same for both users for simplification reasons. As previously stated, we adopt a practical path-loss model with path-loss exponent  $L_P = 2$ . In the simulations, users are randomly distributed in a circle area centered around the BS and of radius  $d = 1000$  m, and are assumed to remain stationary in the cell.

##### B. Closed-Form Expressions Validation

Fig. 2 plots the EC of the NOMA paired users, i.e.,  $C_u$  and  $C_t$  of the strong user and weak user, versus the transmit SNR ( $\rho$ ). The curves are generated based on the derived closed-form formulae shown in (17), (23), (24) and (18). In order to validate the accuracy of the EC results of the considered system (NOMA with FBL), Monte-Carlo simulations are also performed, referred to as MC in the figure. It is observed that the analytical-based results align well with the Monte-Carlo results, which demonstrates the correctness of our analytical derivations. The slight difference between the analysis and simulation findings can be explained by the fact that we relied on some approximations to derive the closed-form expressions. Also, the values of the bivariate Meijer-G function are obtained based on the implementation of the Mellin-Barnes integral representation of the said function in Matlab. As observed, the users' achievable ECs saturate at different values of SNR.

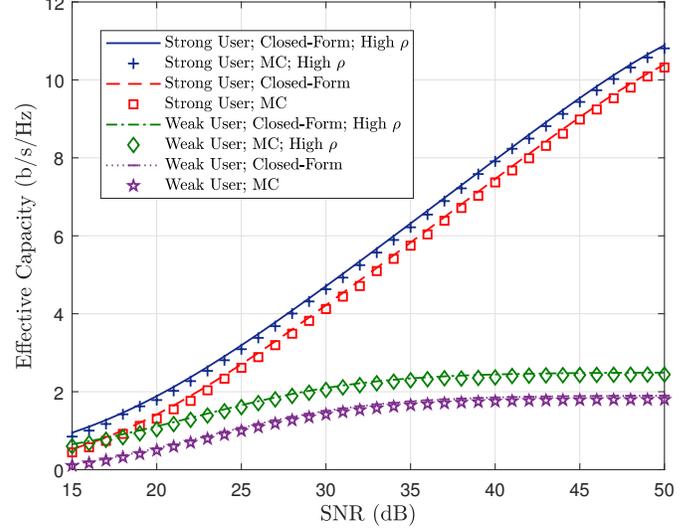


Figure 2. Effective capacity versus transmit SNR for both users.

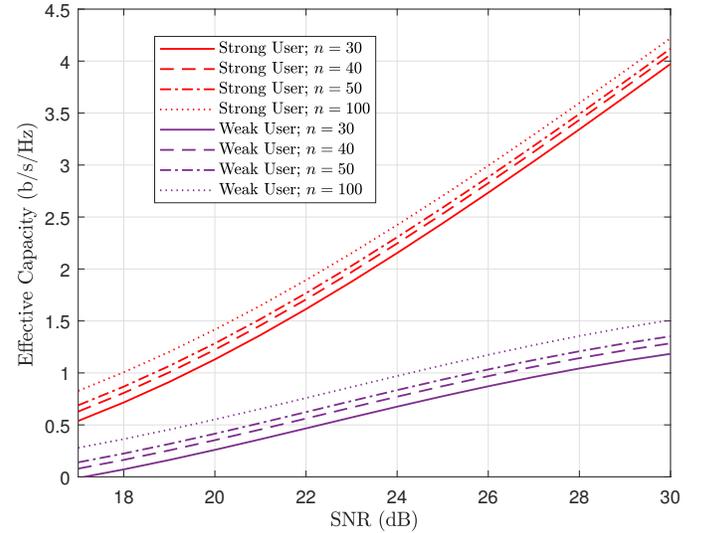


Figure 3. Effective capacity versus SNR for different values of  $n$ .

##### C. Effect of the Main System Parameters on EC

In Fig. 3, the effect of the transmission blocklength  $n$  on the EC of the two-user NOMA is investigated. For both users, increasing  $n$  enhances the EC because longer blocklengths allow for better error correction and more reliable communication. For instance, at low blocklength values (e.g.,  $n = 30$ ), the EC is lower due to higher decoding errors and finite blocklength penalties, which are more pronounced at lower SNR values. This trade-off between reliability and latency is critical for sensitive applications such as extreme-uRLLC, where short blocklengths ensure low delay but reduce capacity, while longer blocklengths improve throughput but at the cost of high latency. The findings emphasize the importance of optimizing  $n$  based on application-specific latency and reliability requirements.

In Fig. 4, the achievable EC of the strong and weak users

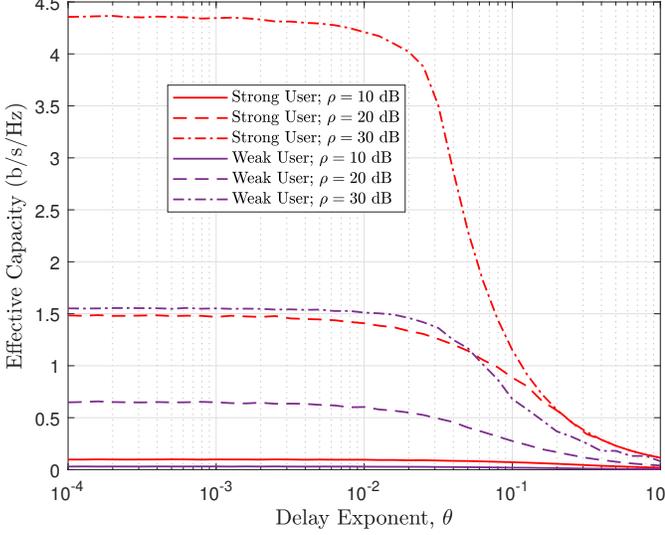


Figure 4. Effective capacity of NOMA strong and weak users versus delay exponent  $\theta$ .

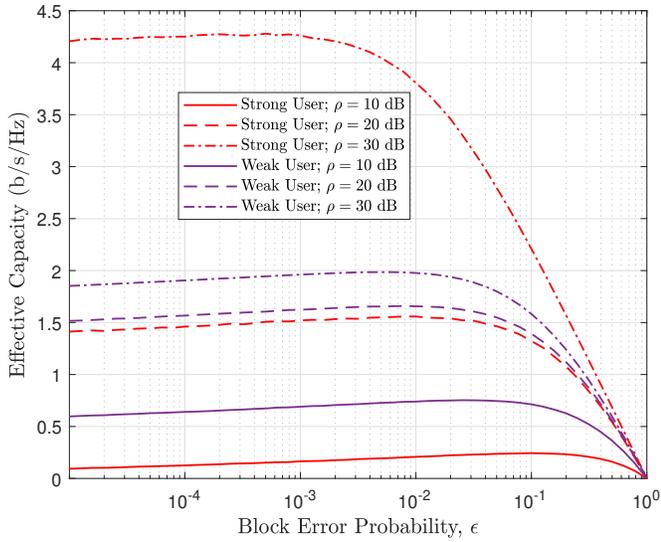


Figure 5. Effective capacity versus the block error probability when  $n = 100$  and  $\theta = 0.01$ .

is plotted versus the delay exponent  $\theta$ , for different transmit SNR values, namely,  $\rho = 10, 20$  and  $30$  dB. It is observed from the curves that raising the delay exponent  $\theta$ , indicating stricter latency requirements, reduces the achievable EC of both users. This is because higher values of  $\theta$  require shorter transmission durations to meet the latency constraints, which may limit the amount of information that can be reliably transmitted within each block. In particular, for fixed values of  $\rho$ , we notice that for low loose delay requirements, i.e., low values of  $\theta$ , the gains in the EC of the strong user are more significant, with a large gap when compared to the gains of the weak user. However, the impact of  $\theta$  on the weak-user's achievable EC may be more pronounced, especially if the weak user operates in higher SNR regimes. The weak-user's achievable EC exhibits a steeper decline with increasing

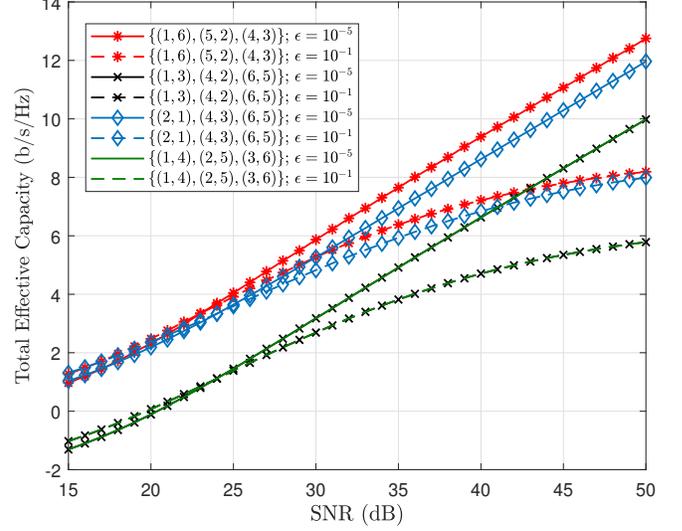


Figure 6. Total effective capacity of multiple NOMA pairs versus transmit SNR, with  $V = 6$ .

$\theta$  compared to the strong user, particularly at higher values of  $\theta$  where stringent latency constraints significantly limit the transmission time and reliability.

In order to further understand how the reliability of the transmission, characterized by  $\epsilon$ , which is one of the most important metrics in FBL regime, impacts the EC of the system, next we analyze the impact of the block error probability on the EC for a two-user NOMA system in FBL.

Fig. 5 plots the EC of the strong and weak users versus the block error probability ( $\epsilon$ ) for various values of SNR ( $\rho$ ). It is observed that when  $\epsilon$  is very small, the transmission is highly reliable. In this regime, the achievable EC tends to approach the channel capacity, as the system can reliably transmit information at rates close to the channel capacity without significant loss. Both users can achieve high EC values when the error probability is extremely low, assuming sufficient power allocation and appropriate decoding techniques. As  $\epsilon$  increases, due to FBL effects or deteriorating channel conditions, the achievable EC starts to decrease. The effects of FBL become more pronounced, leading to higher error probabilities for a given blocklength. The decrease in EC is more significant for users operating at lower SNRs or experiencing more severe FBL effects, i.e., the weak user. This can be analytically explained by the fact that the term  $(1 - \epsilon_{k_1})(1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1} \sqrt{\delta_{k_1}}}$  of the EC formulation shown in (29) is big as compared to  $\epsilon_{k_1}$ . Compared to the strong user, however, in the weak-user's achievable EC formulation,  $\epsilon$  remains more dominant as compared to the term  $(1 - \epsilon_{k_2}) \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} e^{\beta_{k_2} \sqrt{\delta_{k_2}}}$ . For further analysis on the behavior, cf. Eq. (38).

Fig. 6 plots the total rate  $T_{ec}$  versus the transmit SNR of multiple NOMA pairs with short-packet communication. Various sets of users have been paired together depending on their channel conditions. The total number of users is  $V = 6$ , and the power coefficients allocated to the strong and weak users in a NOMA pair are set as  $\alpha_{k_1} = 0.2$  and

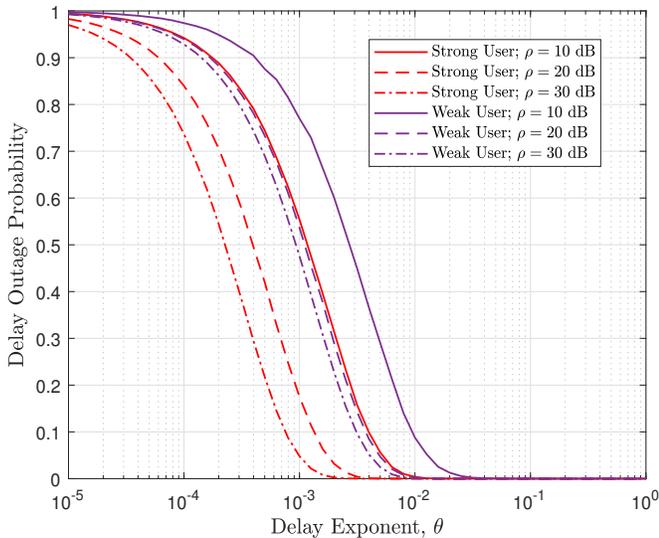


Figure 7. Delay outage probability versus delay exponent when  $n = 100$  and  $\epsilon = 10^{-5}$ .

$\alpha_{k_2} = 0.8$  in all the groups. Here, the QoS delay exponent is assumed to be the same for all users:  $\theta = 0.01$ . It is shown that the best pairing that provides higher effective rates is  $\phi = \{(1, 6), (2, 5), (3, 4)\}$  (cf. red curves in the figure). This demonstrates that when the users with distinct channel conditions are paired together, they can achieve higher  $T_{ec}$  as compared to the pairing of users with less distinct channel conditions. This can be explained by the fact that the effects of interference are decreased when the strongest user and the weakest user are coupled together during the decoding phase, resulting in higher rate gains. It is also noted that the total effective rate drops as  $\epsilon$  increases, as expected.

To understand how different values of  $\theta$  impact the system's ability to meet latency requirements and the occurrence of delay outages, Fig. 7 plots the delay outage probability versus the delay exponent for the strong user and the weak user in a two-user NOMA system. For low values of  $\theta$ , indicating less stringent latency requirements, the system prioritizes latency over reliability. This is due to the fact that  $\epsilon_{k_1}$  is more prominent than the term  $(1 - \epsilon_{k_1})(1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1}\sqrt{\delta_{k_1}}}$  in the EC formulation. Furthermore, at high values of  $\theta$ ,  $(1 - \epsilon_{k_1})(1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1}\sqrt{\delta_{k_1}}}$  becomes very small and, hence,  $\epsilon_{k_1}$  becomes the dominant factor. This means that the system aims to minimize the transmission duration at the expense of reduced reliability. The delay outage probability tends to increase as  $\theta$  approaches zero, since the system imposes strict latency constraints, limiting the transmission time available for each block. Therefore, the delay outage probability is higher for lower values of  $\theta$ . The delay outage probability decreases as  $\theta$  increases. This is because higher values of  $\theta$  relax the latency constraints, allowing for longer transmission duration and, hence, better reliability. Finally, at very high values of  $\theta$  (close to 1), the delay outage probability approaches 0.

In Fig. 8, we plot the delay outage probability versus  $\epsilon$ , for various values of the delay exponent  $\theta$ , of a two-

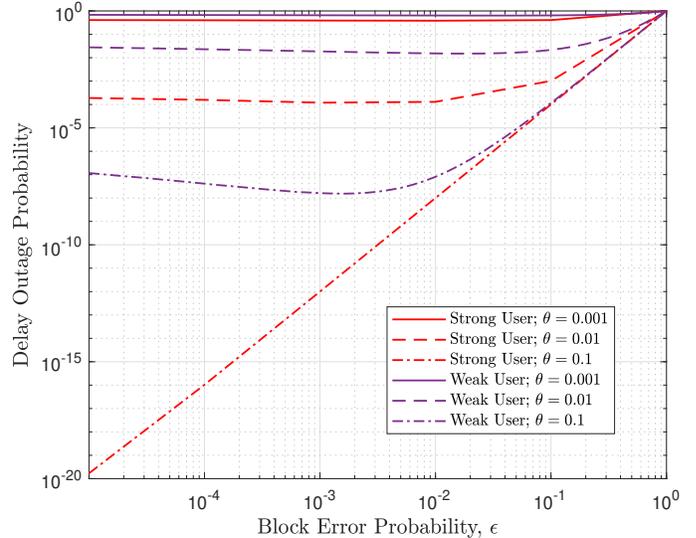


Figure 8. Delay outage probability versus block error probability when  $\rho = 20$  and  $n = 100$ .

user NOMA system. Results are shown for the strong and weak users. At very low error probabilities ( $\epsilon$  close to 0), the transmission is highly reliable. In this regime, the delay outage probability tends to be low since the likelihood of experiencing transmission errors that lead to delays is minimal. Both users in the NOMA system can achieve low delay outage probabilities when the error probability is extremely low. As  $\epsilon$  increases, due to FBL effects or deteriorating channel conditions, the delay outage probability starts to rise. At very high error probabilities ( $\epsilon$  close to 1), the delay outage probability approaches unity for both users and for the different  $\theta$  values considered. This occurs when the error probability becomes so high that the transmissions consistently fail to meet the specified latency constraint, resulting in an outage event in nearly every transmission attempt. In conclusion, our findings confirm the impact of the short-packet communication on the delay outage probability of the NOMA strong and weak users. Furthermore, the user with weak channel conditions does not exhibit significant enhancement in terms of delay outage probability compared to the user with strong channel conditions.

## V. CONCLUSION

We investigated the performance of NOMA in finite block-length (FBL) regime based on the effective capacity (EC) framework. Two scenarios were considered; specifically, the multiple NOMA pairs and the two-user NOMA. Overall reliability needs were assessed by taking the delay outage probability and the block error probability into consideration. In particular, closed-form expressions for the individual EC of the strong and weak users were derived, and Monte-Carlo simulations were leveraged to confirm their accuracy. Our findings showed that the EC decreases with the delay exponent, the block error probability and the blocklength. Furthermore, the total EC of multiple NOMA pairs in FBL regime were investigated, and showed that users experiencing

greater variation in channel conditions achieve higher EC as compared to the users with less distinct channel conditions.

#### APPENDIX A

Here, the closed-form expression for the achievable EC of the strong user ( $k_1$ ) is determined. To begin, the achievable EC is provided based on (9), and is defined as

$$C_{k_1} = -\frac{\ln\left(\mathbb{E}\left[\epsilon_{k_1} + (1 - \epsilon_{k_1})(1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} e^{\beta_{k_1}\sqrt{\delta_{k_1}}}\right]\right)}{\theta_{k_1}n}, \quad (29)$$

where  $\zeta_{k_1} = -\frac{\theta_{k_1}n}{2\ln 2}$  and  $\beta_{k_1} = \theta\sqrt{n}Q^{-1}(\epsilon_{k_1})$ .

To get a closed-form expression for (29), the order statistics PDF from (11) is applied, and converting  $e^{\beta_{k_1}\sqrt{\delta_{k_1}}}$  into a fractional form using the Maclaurin series expansion,<sup>2</sup> such that  $e^{\beta_{k_1}\sqrt{\delta_{k_1}}} \approx 1 + \beta_{k_1}\sqrt{\delta_{k_1}} + \frac{(\beta_{k_1}\sqrt{\delta_{k_1}})^2}{2}$ , the achievable EC of the strong user can be expressed as where  $\sum_p = \sum_{\text{all permutations } V!}$ .

After inserting  $\delta_{k_1} = 1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}$  into Eq. (30), the achievable EC of the strong user reduces to the expression shown in Eq. (31). At this point, by replacing (14) in (31), the achievable EC of the strong user is expressed as follows:

$$C_{k_1} \approx \frac{-1}{\theta_{k_1}n} \ln\left(\int_0^\infty \left(\epsilon_{k_1} + (1 - \epsilon_{k_1}) \times \left( (1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} + \beta_{k_1}(1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} \times \sqrt{1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}} + \frac{\beta_{k_1}^2 \left(\sqrt{1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}}\right)^2}{2} \right) \times \frac{\sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1}}{(k_1 - 1)!(V - k_1)!\rho_{k_1}} \right)\right). \quad (32)$$

With further manipulations, the achievable EC with order statistics from (32) is given by

$$C_{k_1} = \frac{-1}{\theta_{k_1}n} \ln\left(\frac{1}{(k_1 - 1)!(V - k_1)!\rho_{k_1}} \int_0^\infty \left(\epsilon_{k_1} + (1 - \epsilon_{k_1}) \left( (1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} \left( 1 + \beta_{k_1}\sqrt{1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}} + \frac{\beta_{k_1}^2 \left(1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}\right)^2}{2} \right) \times \sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1} \right)\right)\right).$$

<sup>2</sup>This approximation is commonly used when dealing with small values of  $\beta_{k_1}\sqrt{\delta_{k_1}}$ , and it simplifies the integration or computation of the expression.

Now, simplifying the expression further, we get

$$C_{k_1} = -\frac{1}{\theta_{k_1}n} \ln\left(\frac{1}{(k_1 - 1)!(V - k_1)!\rho_{k_1}} \times \left( I_1 + (1 - \epsilon_{k_1}) \left( I_2 + \beta_{k_1}I_3 + \frac{\beta_{k_1}^2}{2}I_4 \right) \right)\right), \quad (33)$$

where  $I_1$ ,  $I_2$ ,  $I_3$  and  $I_4$  are expressed as follows:

$$\begin{aligned} I_1 &= \int_0^\infty \epsilon_{k_1} \sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1} \\ &\approx \epsilon_{k_1} \sum_p \int_0^\infty \left(\frac{\gamma_{k_1}^{k_1-1}}{\prod_{a=1}^{k_1-1} \rho_a}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1} \\ &\approx \epsilon_{k_1} \sum_p \frac{\Gamma(k_1)}{\prod_{a=1}^{k_1-1} \rho_a \left(\sum_{c=k_1+1}^V \left(\frac{1}{\rho_c}\right) + \frac{1}{\rho_{k_1}}\right)^{k_1+1}}, \end{aligned} \quad (34)$$

$$\begin{aligned} I_2 &= \int_0^\infty (1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} \sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\nu_1} d\gamma_{k_1} \\ &= \sum_p \frac{\Gamma(k_1)\alpha_{k_1}^{k_1}}{\prod_{a=1}^{k_1-1} \rho_a} \mathbf{H}(k_1, a_{11}, \alpha_{k_1}\nu_1), \end{aligned} \quad (35)$$

where  $\nu_1$  and  $a_{11}$  are as defined after Eq. (17),

$$\begin{aligned} I_3 &= \int_0^\infty (1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} \sqrt{1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}} \\ &\quad \times \sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1} \\ &= \sum_p \frac{\alpha_{k_1}^{k_1}\Gamma(k_1)}{\prod_{a=1}^{k_1-1} \rho_a} \left(\mathbf{H}(u, a_{11}, \alpha_{k_1}\nu_1) - \frac{1}{2}\mathbf{H}(k_1, a_{12}, \alpha_{k_1}\nu_1)\right), \end{aligned} \quad (36)$$

where  $a_{11}$  and  $a_{12}$  are as defined after Eq. (17), and

$$\begin{aligned} I_4 &= \int_0^\infty (1 + \alpha_{k_1}\gamma_{k_1})^{2\zeta_{k_1}} \left(1 - (1 + \alpha_{k_1}\gamma_{k_1})^{-2}\right) \\ &\quad \times \sum_p \prod_{a=1}^{k_1-1} \left(1 - e^{-\frac{\gamma_{k_1}}{\rho_a}}\right) e^{-\sum_{c=k_1+1}^V \frac{\gamma_{k_1}}{\rho_c} - \frac{\gamma_{k_1}}{\rho_{k_1}}} d\gamma_{k_1} \\ &= \sum_p \frac{\alpha_{k_1}^{k_1}\Gamma(k_1)}{\prod_{a=1}^{k_1-1} \rho_a} \left(\mathbf{H}(k_1, a_{11}, \alpha_{k_1}\nu_1) - \mathbf{H}(k_1, a_{12}, \alpha_{k_1}\nu_1)\right). \end{aligned} \quad (37)$$

Finally, the closed-form expression for the strong-user's achievable EC is expressed as shown in (17). By following the same steps as in the above, the closed-form expression for the strong user's achievable EC at high SNR can also be derived as per (23).

#### APPENDIX B

The achievable EC of the weak user and the corresponding closed-form expression are derived using the same techniques

$$C_{k_1} \approx -\frac{1}{\theta_{k_1} n} \ln \left( \int_0^\infty \left( \epsilon_{k_1} + (1 - \epsilon_{k_1})(1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} \left( 1 + \beta_{k_1} \sqrt{\delta_{k_1}} + \frac{(\beta_{k_1} \sqrt{\delta_{k_1}})^2}{2} \right) \right) \right. \\ \left. \times \frac{1}{(k_1 - 1)! (V - k_1)!} \sum_p F_{j_1}(\gamma_{k_1}) \cdots F_{j_{k_1-1}}(\gamma_{k_1}) f_{j_{k_1}}(\gamma_{k_1}) (1 - F_{j_{k_1+1}}(\gamma_{k_1})) \cdots (1 - F_{j_V}(\gamma_{k_1})) d\gamma_{k_1} \right), \quad (30)$$

$$C_{k_1} \approx \frac{-1}{\theta_{k_1} n} \ln \left( \int_0^\infty \left( \epsilon_{k_1} + (1 - \epsilon_{k_1}) \right. \right. \\ \left. \left. \times \left( (1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} + \beta_{k_1} (1 + \alpha_{k_1} \gamma_{k_1})^{2\zeta_{k_1}} \sqrt{1 - (1 + \alpha_{k_1} \gamma_{k_1})^{-2}} + \frac{\beta_{k_1}^2 \left( \sqrt{1 - (1 + \alpha_{k_1} \gamma_{k_1})^{-2}} \right)^2}{2} \right) \right. \right. \\ \left. \left. \times \frac{\sum_p}{(k_1 - 1)! (V - k_1)!} F_{j_1}(\gamma_{k_1}) \cdots F_{j_{k_1-1}}(\gamma_{k_1}) f_{j_{k_1}}(\gamma_{k_1}) (1 - F_{j_{k_1+1}}(\gamma_{k_1})) \cdots (1 - F_{j_V}(\gamma_{k_1})) d\gamma_{k_1} \right) \right). \quad (31)$$

as the one used for the strong user. Based on (9), the weak user's achievable EC is stated as

$$C_{k_2} = -\frac{1}{\theta_{k_2} n} \\ \times \ln \left( \mathbb{E} \left[ \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} e^{\beta_{k_2} \sqrt{\delta_{k_2}}} \right] \right), \quad (38)$$

where  $\zeta_{k_2} = -\frac{\theta_{k_2} n}{2 \ln 2}$  and  $\beta_{k_2} = \theta \sqrt{n} Q^{-1}(\epsilon_{k_2})$ .

To find the closed-form formula, the above expression can be simplified by replacing the order statistics from (11) into (38). As such, the above equation reduces to Eq. (39).

Using the Maclaurin series expansion,  $e^{\beta_{k_2} \sqrt{\delta_{k_2}}} \approx 1 + \beta_{k_2} \sqrt{\delta_{k_2}} + \frac{(\beta_{k_2} \sqrt{\delta_{k_2}})^2}{2}$ , and expanding the order statistics from (11), and then replacing them in (39), the achievable EC of user  $k_2$  reduces to Eq. (40).

Following that, by inserting  $\delta_{k_2} = 1 - \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2}$  in (40), and making use of (14) into (40), the weak user's achievable EC is reformulated as shown in (41), in which  $\sum_p = \sum_{\text{all permutations } V!}$ .

The achievable EC of the weak user is formulated as shown in (42). To simplify (42), we use the generalized binomial expansion [47] and expand the terms  $\left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}}$  such that

$$\left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} = \left( \frac{1}{\alpha_{k_1}} \right)^{2\zeta_{k_2}} \left( 1 + \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}}. \quad (43)$$

Then, the term  $\left( 1 + \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}}$  can be expanded as

$$\left( 1 + \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} = \sum_{s=0}^{\infty} \binom{2\zeta_{k_2}}{s} \left( \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^s, \quad (44)$$

where, from [47], it is clear that

$$(1 + a)^x = \sum_{y=0}^{\infty} \binom{x}{y} a^y \quad \text{for } |a| < 1. \quad (45)$$

For  $y \geq 1$ ,  $\binom{x}{y}$  can be written as

$$\binom{x}{y} = \frac{x(x-1) \cdots (x-y+1)}{y!} = \frac{(x)_y}{y!}, \quad (46)$$

where  $\binom{x}{0} = 1$ , and  $(\cdot)_y$  is the Pochhammer symbol.

Using the first-order binomial expansion of  $\left( 1 - e^{-\frac{\gamma_t}{\rho_t}} \right)$ , (42) can be expressed as shown in (47), in which  $I_{12}$ ,  $I_{13}$ , and  $I_{14}$  are given in (48), (49) and (50), with  $\sum_p = \sum_{\text{all permutations } V!}$  and  $\nu_2 = \sum_{c=k_2+1}^V \left( \frac{\gamma_{k_2}}{\rho_c} \right) - \frac{\gamma_{k_2}}{\rho_{k_2}}$ .

Finally, we reach the closed-form expression for  $C_t$  as shown in (18). The closed-form expression for the achievable EC of the weak user at high SNR presented in (24) can also be derived by following the above steps.

## REFERENCES

- [1] V. Quy, A. Chehri, N. Quy, N. Han, and N. Ban, "Innovative trends in the 6G era: A comprehensive survey of architecture, applications, technologies, and challenges," *IEEE Access*, vol. 11, pp. 39 824–39 844, Apr. 2023.
- [2] Y. Zhang, W. Cheng, and W. Zhang, "Multiple access integrated adaptive finite blocklength for ultra-low delay in 6g wireless networks," *IEEE Trans. Wireless Commun.*, pp. 1–1, July 2023.
- [3] C. Wang *et al.*, "On the road to 6G: Visions, requirements, key technologies, and testbeds," *IEEE Commun. Surv. Tuts.*, vol. 25, no. 2, pp. 905–974, Feb. 2023.
- [4] A. Dogra, R. Jha, and S. Jain, "A survey on beyond 5G network with the advent of 6G: architecture and emerging technologies," *IEEE Access*, vol. 9, pp. 67 512–67 547, Oct. 2021.
- [5] Y. Xu, Y. Mao, O. Dizdar, and B. Clerckx, "Rate-splitting multiple access with finite blocklength for short-packet and low-latency downlink communications," *IEEE Trans. Veh. Techno.*, vol. 71, no. 11, pp. 12 333–12 337, Nov. 2022.
- [6] X. Zhang, G. Han, D. Zhang, and B. Shim, "Sparse vector coding-based superimposed transmission for short packet urllc," *Apr. 2021*.
- [7] M. Siddiqui, H. Abumarshoud, L. Bariah, S. Muhaidat, M. Imran, and L. Mohjazi, "URLLC in beyond 5G and 6G networks: An interference management perspective," *IEEE Access*, vol. 11, pp. 54 639–54 663, June 2023.
- [8] M. Amjad, L. Musavian, and S. Aissa, "NOMA versus OMA in finite blocklength regime: Link-layer rate performance," *IEEE Trans. Veh. Techno.*, vol. 69, no. 12, pp. 16 253–16 257, Dec. 2020.
- [9] Y. Chen, H. Lu, L. Qin, Y. Deng, and A. Nallanathan, "When xURLLC meets NOMA: A stochastic network calculus perspective," *IEEE Commun. Mag.*, pp. 1–7, Dec. 2023.

$$C_{k_2} = -\frac{1}{\theta_{k_2} n} \times \ln \left( \int_0^\infty \left( \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} e^{\beta_{k_2} \sqrt{\delta_{k_2}}} \right) f_{k_2:V}(\gamma_{k_2}) d\gamma_{k_2} \right). \quad (39)$$

$$C_{k_2} \approx -\frac{1}{\theta_{k_2} n} \ln \left( \int_0^\infty \left( \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} \left( 1 + \beta_{k_2} \sqrt{\delta_{k_2}} + \frac{(\beta_{k_2} \sqrt{\delta_{k_2}})^2}{2} \right) \right) \frac{1}{(k_2 - 1)! (V - k_2)!} \right. \\ \left. \times \sum_p F_{j_1}(\gamma_{k_2}) \cdots F_{j_{k_2-1}}(\gamma_{k_2}) f_{j_{k_2}}(\gamma_{k_2}) (1 - F_{j_{k_2+1}}(\gamma_{k_2})) \cdots (1 - F_{j_V}(\gamma_{k_2})) d\gamma_{k_2} \right). \quad (40)$$

$$C_{k_2} \approx \frac{-1}{\theta_{k_2} n} \ln \left( \frac{1}{(k_2 - 1)! (V - k_2)!} \left( \sum_p \int_0^\infty \left( \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} \right. \right. \right. \right. \\ \left. \left. \left. + \beta_{k_2} \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\zeta_{k_2}} \sqrt{1 + \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2}} + \frac{\beta_{k_2}^2 \left( 1 + \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2} \right)}{2} \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\xi_{k_2}} \right) \right) \right. \\ \left. \times \prod_{a=1}^{i-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) \frac{1}{\rho_i} e^{-\frac{\gamma}{\rho_i}} \prod_{c=i+1}^V e^{-\frac{\gamma_{k_2}}{\rho_c}} d\gamma_{k_2} \right). \quad (41)$$

$$C_{k_2} = \frac{-1}{\theta_{k_2} n} \ln \left( \frac{1}{(k_2 - 1)! (V - k_2)! \rho_{k_2}} \left( \sum_p \int_0^\infty \left( \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\xi_{k_2}} + \beta_{k_2} \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\xi_{k_2}} \right. \right. \right. \right. \\ \left. \left. \left. \times \sqrt{1 + \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2}} + \frac{\left( \beta_{k_2} \sqrt{1 + \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2}} \right)^2 \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{2\xi_{k_2}}}{2} \right) \prod_{a=1}^{k_2-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) e^{-\sum_{c=k_2+1}^V \left( \frac{\gamma_{k_2}}{\rho_c} - \frac{\gamma_{k_2}}{\rho_{k_2}} \right)} d\gamma_{k_2} \right) \right). \quad (42)$$

$$C_{k_2} = -\frac{1}{\theta_{k_2} n} \ln \left( \frac{1}{(k_2 - 1)! (V - k_2)! \rho_{k_2}} \sum_p \left( \epsilon_{k_2} + (1 - \epsilon_{k_2}) \left( \alpha_{k_1}^{-2\zeta_{k_2}} \left( \right. \right. \right. \right. \\ \left. \left. \left. \times \int_0^\infty \underbrace{\left( 1 + 2\zeta_{k_2} \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left( \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^s \right)}_{I_{12}} \prod_{a=1}^{k_2-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) e^{-\nu_2} d\gamma_{k_2} \right. \right. \right. \\ \left. \left. \left. + \beta_{k_2} \int_0^\infty \underbrace{\sqrt{1 - \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2}} \left( 1 + 2\zeta_{k_2} \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left( \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^s \right)}_{I_{13}} \prod_{a=1}^{k_2-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) e^{-\nu_2} d\gamma_{k_2} \right. \right. \right. \\ \left. \left. \left. + \frac{\beta_{k_2}^2}{2} \right. \right. \right. \\ \left. \left. \left. \times \int_0^\infty \underbrace{\left( 1 - \left( \frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^{-2} \right) \left( 1 + 2\zeta_{k_2} \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left( \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^s \right)}_{I_{14}} \prod_{a=1}^{k_2-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) e^{-\nu_2} d\gamma_{k_2} \right) \right) \right). \quad (47)$$

$$I_{12} = \int_0^\infty \left( 1 + 2\zeta_{k_2} \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left( \frac{1}{\alpha_{k_1}} \right)^{2\zeta_{k_2}} \left( \frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1} \right)^s \right) \prod_{a=1}^{k_2-1} \left( 1 - e^{-\frac{\gamma_{k_2}}{\rho_a}} \right) e^{-\nu_2} d\gamma_{k_2} \\ = \frac{\Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a (\nu_2)^{k_2+1}} + \frac{2\zeta_{k_2} (\alpha_{k_1} - 1)}{\prod_{a=1}^{k_2-1} \rho_a} \alpha_{k_1}^{-k_2} e^{\frac{\nu_2}{\alpha_{k_1}}} \Gamma(k_2) \Gamma(1 - k_2, \frac{\nu_2}{\alpha_{k_1}}) + \frac{\sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s}{\prod_{a=1}^{k_2-1} \rho_a} \\ \times \left( \frac{\alpha_{k_1}}{\Gamma(s)} \Gamma(k_2) \Gamma(s - k_2) {}_1F_1(k_2, k_2 - s + 1, \frac{\nu_2}{\alpha_{k_1}}) + \alpha_{k_1}^{-s} \nu_2^{s-k_2} \Gamma(k_2 - s) {}_1F_1(s, 1 - k_2 + s, \frac{\nu_2}{\alpha_{k_1}}) \right). \quad (48)$$

$$\begin{aligned}
I_{13} &= \int_0^\infty \sqrt{1 + \left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{-2}} \left(1 + \frac{2\zeta_{k_2}(\alpha_{k_1} - 1)}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left(\frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^s\right) \prod_{a=1}^{k_2-1} \left(1 - e^{-\frac{\gamma_{k_2}}{\rho_a}}\right) e^{-\nu_2} d\gamma_{k_2} \\
&= \frac{\Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a \nu_2^{k_2+1}} + \frac{2\zeta_{k_2}(\alpha_{k_1} - 1) \alpha_{k_1}^{-k_2} e^{\frac{\nu_2}{\alpha_{k_1}} \Gamma(k_2) \Gamma(1-k_2, \frac{\nu_2}{\alpha_{k_1}})}}{\prod_{a=1}^{k_2-1} \rho_a} + \frac{\sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}}}{\prod_{a=1}^{k_2-1} \rho_a} \\
&\times \left( \frac{\alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma(s-k_2)}{\Gamma(s)} {}_1F_1\left(k_2, k_2 - s + 1, \frac{\nu_2}{\alpha_{k_1}}\right) + \alpha_{k_1}^{-s} (\nu_2)^{s-k_2} \Gamma(k_2 - s) {}_1F_1\left(s, 1 - k_2 + s, \frac{\nu_2}{\alpha_{k_1}}\right) \right) \\
&+ \frac{1}{2} \left( \Gamma(k_2) \nu_2^{-k_2} \left(1 - (\alpha_{k_1} - 1)^2 t + (\alpha_{k_1} - 1) e^{\nu_2} k_2 \left(1 - \nu_2 - k_2 + \alpha_{k_1} (1 + \nu_2 + k_2)\right)\right) \right) \\
&+ \frac{2\zeta_{k_2}(\alpha_{k_1} - 1)}{\prod_{a=1}^{k_2-1} \rho_a} \left( \nu_2^{1-k_2} \left(1 - \alpha_{k_2} + e^{\nu_2} \left(1 + (-1 + \alpha_{k_1}) \nu_2 + (-1 + \alpha_{k_1}) t\right) \text{Ei}(k_2, \nu_2)\right) \Gamma(k_2) \right) \\
&+ \frac{\sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left(\frac{\nu_2}{2}\right)^{-k_2}}{\prod_{a=1}^{k_2-1} \rho_a \sqrt{\pi} \Gamma(-2+s)} G_{2,0:1,1:1,1}^{2,0:1,1:1,1} \left( \begin{matrix} 1 - k_2, 1 - k_2 - \frac{1}{2} \\ 1 - k_2, \end{matrix} \middle| \begin{matrix} -1 \\ 0 \end{matrix} \middle| \begin{matrix} 3 - s \\ 0 \end{matrix} \middle| \begin{matrix} \frac{2}{\nu_2}, \frac{2\alpha_{k_1}}{\nu_2} \end{matrix} \right).
\end{aligned} \tag{49}$$

$$\begin{aligned}
I_{14} &= \int_0^\infty \left(1 - \left(\frac{\gamma_{k_2} + 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^{-2}\right) \left(1 + \frac{2\zeta_{k_2}(\alpha_{k_1} - 1)}{\alpha_{k_1} \gamma_{k_2} + 1} + \sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left(\frac{\alpha_{k_1} - 1}{\alpha_{k_1} \gamma_{k_2} + 1}\right)^s\right) \prod_{a=1}^{k_2-1} \left(1 - e^{-\frac{\gamma_{k_2}}{\rho_a}}\right) e^{-\nu_2} d\gamma_{k_2} \\
&= \frac{\Gamma(k_2)}{\prod_{a=1}^{k_2-1} \rho_a \nu_2^{k_2+1}} + \frac{2\zeta_{k_2}(\alpha_{k_1} - 1) \alpha_{k_1}^{-k_2} e^{\frac{\nu_2}{\alpha_{k_1}} \Gamma(k_2) \Gamma(1-k_2, \frac{\nu_2}{\alpha_{k_1}})}}{\prod_{a=1}^{k_2-1} \rho_a} \\
&+ \frac{\sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left( \alpha_{k_1}^{-k_2} \Gamma(k_2) \Gamma(s-k_2) {}_1F_1\left(k_2, k_2 - s + 1, \frac{\nu_2}{\alpha_{k_1}}\right) + \Gamma(s) \alpha_{k_1}^{-s} \nu_2^{s-k_2} \Gamma(k_2 - s) {}_1F_1\left(s, 1 - k_2 + s, \frac{\nu_2}{\alpha_{k_1}}\right) \right)}{\Gamma(s) \prod_{a=1}^{k_2-1} \rho_a} \\
&- \left( \Gamma(k_2) \nu_2^{-k_2} \left(1 - (\alpha_{k_1} - 1)^2 t + (\alpha_{k_1} - 1) e^{\nu_2} k_2 \left(1 - \nu_2 - k_2 + \alpha_{k_1} (1 + \nu_2 + k_2)\right)\right) \right) \\
&+ \frac{2\zeta_{k_2}(\alpha_{k_1} - 1)}{\prod_{a=1}^{k_2-1} \rho_a} \left( \nu_2^{1-k_2} \left(1 - \alpha_{k_2} + e^{\nu_2} \left(1 + (-1 + \alpha_{k_1}) \nu_2 + (-1 + \alpha_{k_1}) t\right) \text{Ei}(k_2, \nu_2)\right) \Gamma(k_2) \right) \\
&+ \frac{\sum_{s=2}^\infty \binom{2\zeta_{k_2}}{s} (\alpha_{k_1} - 1)^s \left(\frac{1}{\alpha_{k_1}}\right)^{2\zeta_{k_2}} \left(\frac{\nu_2}{2}\right)^{-k_2}}{\prod_{a=1}^{k_2-1} \rho_a \sqrt{\pi} \Gamma(-2+s)} G_{2,0:1,1:1,1}^{2,0:1,1:1,1} \left( \begin{matrix} 1 - k_2, 1 - k_2 - \frac{1}{2} \\ 1 - k_2, \end{matrix} \middle| \begin{matrix} -1 \\ 0 \end{matrix} \middle| \begin{matrix} 3 - s \\ 0 \end{matrix} \middle| \begin{matrix} \frac{2}{\nu_2}, \frac{2\alpha_u}{\nu_2} \end{matrix} \right).
\end{aligned} \tag{50}$$

- [10] Z. Xiang, W. Yang, Y. Cai, Z. Ding, Y. Song, and Y. Zou, "NOMA-assisted secure short-packet communications in IoT," *IEEE Wireless Commun.*, vol. 27, no. 4, pp. 8–15, Aug. 2020.
- [11] Q. Chen, J. W. J. Wang, and H. Jiang, "Coexistence of URLLC and eMBB services in MIMO-NOMA systems," *IEEE Trans. Veh. Technol.*, pp. 1–13, Sep. 2022.
- [12] G. Tran and S. Kim, "Performance evaluation of short packet communications in NOMA VLC systems with imperfect CSI," *IEEE Access*, vol. 10, pp. 49 781–49 793, May 2022.
- [13] F. Nadeem, M. Shirvanimoghaddam, Y. Li, and B. Vucetic, "Delay-sensitive NOMA-HARQ for short packet communications," *MDPI Entropy (Basel)*, July 2021.
- [14] H. Ji, S. Park, and B. Shim, "Sparse vector coding for ultra reliable and low latency communications," *IEEE Transactions on Wireless Communications*, vol. 17, pp. 6693–6706, 2017. [Online]. Available: <https://api.semanticscholar.org/CorpusID:52962754>
- [15] B. Adhikari, M. Jaseemuddin, and A. Anpalagan, "Resource allocation for co-existence of eMBB and URLLC services in 6G wireless networks: A survey," *IEEE Access*, vol. 12, pp. 552–581, Dec. 2023.
- [16] X. Sun, S. Yan, N. Yang, Z. Ding, C. Shen, and Z. Zhong, "Short-packet downlink transmission with non-orthogonal multiple access," *IEEE Trans. Wireless Commun.*, vol. 17, no. 7, pp. 4550–4564, July 2018.
- [17] Z. Ding, Z. Yang, P. Fan, and H. Poor, "On the performance of non-orthogonal multiple access in 5G systems with randomly deployed users," *IEEE Sig. Process. Letts.*, vol. 21, no. 12, pp. 1501–1505, Dec. 2014.
- [18] Y. Liu, Z. Qin, M. Elkashlan, Z. Ding, A. Nallanathan, and L. Hanzo, "Nonorthogonal multiple access for 5G and beyond," *Proc. IEEE*, vol. 105, no. 12, pp. 2347–2381, Dec. 2017.
- [19] X. Wang, J. Wang, L. He, and J. Song, "Outage analysis for downlink NOMA with statistical channel state information," *IEEE Wireless Commun. Letts.*, vol. 7, no. 2, pp. 142–145, Apr. 2018.
- [20] S. Herfiah and Iskandar, "Effect of imperfect SIC in non-orthogonal multiple access (NOMA) over rayleigh fading channels," in *Proc. Int. Conf. Wireless Telema. (ICWT)*, Aug. 2021, pp. 1–4.
- [21] D. Do and T. Nguyen, "Impacts of imperfect SIC and imperfect hardware in performance analysis on AF non-orthogonal multiple access network," *Telecommun. Sys.: Modelling, Analysis, Design and Management*, vol. 72, no. 4, pp. 579–593, Dec. 2019.
- [22] J. Zhao, X. Yue, S. Kang, and W. Tang, "Joint effects of imperfect CSI and SIC on NOMA based satellite-terrestrial systems," *IEEE Access*, vol. 9, pp. 12 545–12 554, Jan. 2021.
- [23] G. Tran and S. Kim, "Performance analysis of short packets in NOMA VLC systems," *IEEE Access*, vol. 10, pp. 6505–6517, Jan. 2022.
- [24] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.
- [25] Y. Polyanskiy, H. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.
- [26] L. Yuan, Q. Du, and F. Fang, "Performance analysis of full-duplex cooperative NOMA short-packet communications," *IEEE Trans. Veh. Technol.*, vol. 71, no. 12, pp. 13 409–13 414, Dec. 2022.
- [27] X. Zhang, X. Yue, and S. Kang, "Effective capacity analysis of NOMA networks with short packets," *Applied Sciences*, vol. 11, no. 23, Dec. 2021. [Online]. Available: <https://www.mdpi.com/2076-3417/11/23/11438>

- [28] M. Shehab, H. Alves, and M. Latva-aho, "Effective capacity and power allocation for machine-type communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 4, pp. 4098–4102, Apr. 2019.
- [29] M. Amjad and L. Musavian, "Performance analysis of NOMA for ultra-reliable and low-latency communications," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2018, pp. 1–5.
- [30] Y. Yu, H. Chen, Y. Li, Z. Ding, and B. Vucetic, "On the performance of non-orthogonal multiple access in short-packet communications," *IEEE Commun. Letts.*, vol. 22, no. 3, pp. 590–593, Mar. 2018.
- [31] Q. Zhang, Z. Liang, Q. Li, and J. Qin, "Buffer-aided non-orthogonal multiple access relaying systems in rayleigh fading channels," *IEEE Transactions on Communications*, vol. 65, no. 1, pp. 95–106, Jan. 2017.
- [32] W. Yu, L. Musavian, and Q. Ni, "Link-layer capacity of NOMA under statistical delay QoS guarantees," *IEEE Transactions on Communications*, vol. 66, no. 10, pp. 4907–4922, Oct. 2018.
- [33] J. Yao, Q. Zhang, and J. Qin, "Joint decoding in downlink NOMA systems with finite blocklength transmissions for ultrareliable low-latency tasks," *IEEE Int. Things J.*, vol. 9, no. 18, pp. 17 705–17 713, Sep. 2022.
- [34] Y. Zhang, T. Zhong, Y. Wang, J. Wang, K. Zheng, and X. You, "Max-Min fairness for uplink NOMA systems with finite blocklength," *IEEE Trans. Veh. Technol.*, pp. 1–6, Oct. 2023.
- [35] W. Van Assche, *Ordinary special functions*. Oxford: Academic Press, 2006, pp. 637–645.
- [36] G. B. Arfken, H. J. Weber, and F. E. Harris, "Chapter 12 - further topics in analysis," in *Mathematical Methods for Physicists (Seventh Edition)*, seventh edition ed. Boston: Academic Press, 2013, pp. 551–598.
- [37] J. Choi, "An effective capacity-based approach to multi-channel low-latency wireless communications," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2476–2486, Mar. 2019.
- [38] W. Yang, G. Durisi, T. Koch, and Y. Polyanskiy, "Block-fading channels at finite blocklength," in *Proc. Int. Symp. Wireless Commun. Sys. ISWCS*, Aug. 2013, pp. 1–4.
- [39] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Trans. Wireless Commun.*, vol. 2, no. 4, pp. 630–643, July 2003.
- [40] C. Chang, "Stability, queue length, and delay of deterministic and stochastic queueing networks," *IEEE Trans. Autom. Control*, vol. 39, no. 5, pp. 913–931, May 1994.
- [41] —, "Performance guarantees in communication networks." Springer Science & Business Media., 2012.
- [42] R. Léandre, *A simple proof for a large deviation theorem*. Basel: Birkhäuser Basel, 1993, pp. 72–76. [Online]. Available: [https://doi.org/10.1007/978-3-0348-8555-3\\_5](https://doi.org/10.1007/978-3-0348-8555-3_5)
- [43] J. Bucklew, "Introduction to rare event simulation." Springer Science & Business Media., 2013.
- [44] M. GURSOY, "Throughput analysis of buffer-constrained wireless systems in the finite blocklength regime," *EURASIP J. Wireless Commun. Net.*, vol. 2013, no. 1, p. 290, Dec. 2013.
- [45] R. Bapat and M. Beg, "Order statistics for nonidentically distributed variables and permanents," *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 51, no. 1, pp. 79–93, 1989. [Online]. Available: <http://www.jstor.org/stable/25050725>
- [46] A. Jamjoom and Z. Al-Saiary, "Computing the moments of order statistics from independent nonidentically distributed exponentiated frechet variables," *J. of Probability and Statistics*, pp. 1687–952X, Mar. 2012.
- [47] M. Abramowitz, I. Stegun, and R. Romer, "Handbook of mathematical functions with formulas, graphs, and mathematical tables," 1988.



**Zina Mohamed** (Member, IEEE) received the bachelor's degree in electrical engineering from the National Engineering School of Sfax, Tunisia, in 2017, the master's degree in embedded systems from the National Engineering School of Sfax, Tunisia, in cooperation with the Technische Universität Chemnitz, Germany, in 2017, and the Ph.D. degree in telecommunications from the Institut National de la Recherche Scientifique (INRS), Montreal, QC, Canada, in 2023.

She is currently a Postdoctoral Research Fellow at INRS, Canada. Her research interests include reconfigurable intelligent surfaces, wireless power transfer, energy efficiency, optimization, and UAV communications. In these areas, she has published in several reputed journals and conferences of the IEEE. She is recipient of the INRS-Tunisia Ph.D. Scholarship, and a co-recipient of the 2021 International Wireless Communications and Mobile Computing Conference Best Paper Award.



**Muhammad Amjad** (Member, IEEE) is currently a Lecturer in Computer Science at the University of Wolverhampton, UK. He received his M.S. degree in Computer Science from the COMSATS Institute of Information Technology, Islamabad, Pakistan, and his Ph.D. from the School of Computer Science and Electronic Engineering, University of Essex, UK. His research interests include 5G/B5G networks, ultra-reliable low-latency communications (uRLLC), wireless power transfer, and energy harvesting. He is also a Fellow of advance HEA (FHEA). He received the Best Paper Award at the 2016 ACM International Conference on Computing, Communication and Networking Technologies, USA, and was honored with the 2019 Top Peer Reviewer Award in the field of Computer Science by Publons (Clarivate Analytics - Web of Science Group).



**Leila Musavian** (Member, IEEE) is a Professor at University of Essex. Previously, she was Deputy Pro-Vice-Chancellor for Research at University of Essex and Reader in Telecommunications at the School of Computer Science and Electronic Engineering. Prior to that, she was Lecturer/Senior Lecturer at InfoLab21, Lancaster University, Research Associate at McGill University, Canada, Research Associate at Loughborough University, UK, and Post-Doctoral Fellow at INRS (Institut National de la Recherche Scientifique), Montreal, Canada.

Her research interests lie in Radio Resource Management for 6G/B5G communications, low latency communications, Holographic MIMO Communications, NTN, and SWIPT. She is currently editor of the IEEE Communications Surveys and Tutorials (COMST) and has previously been Editor of IEEE TRANSACTIONS OF WIRELESS COMMUNICATIONS. She was the Conference Workshop Co-Chair of VTC-Spring 2020, the Wireless Communications Symposium Leading Co-Chair for IEEE ICC 2021, Canada, conference TPC Co-Chair of IEEE CAMAD 2021, Portugal, and Track 2 Lead chair of IEEE WCNC 2023. She is the founding chair and co-chair of the IEEE UK & Ireland Future Networks local group.



**Sonia Aïssa** (Fellow, IEEE) received her Ph.D. degree in Electrical and Computer Engineering from McGill University, Montreal, QC, Canada, in 1998. Since then, she has been with the Institut National de la Recherche Scientifique (INRS), Montreal, QC, Canada, where she is Cyrille-Duquet Chair Professor. During her career in telecommunications, she has held various research posts in Canada, Japan, Malaysia, Turkey, and the UK. Her research interests include the modeling, design, performance analysis and optimization of wireless systems and networks.

Prof. Aïssa is Fellow of the IEEE and Fellow of the Canadian Academy of Engineering. Her awards include the NSERC University Faculty Award 1999; the Quebec Government FRQNT Strategic Faculty Fellowship 2001-2006; the INRS Performance Award multiple times, for outstanding achievements in research, teaching and outreach; the 2007 FRQNT-SYTACom Technical Community Service Award; the 2021 IEEE WICE Outstanding Achievement Award, and the 2022 IEEE VTS Women's Distinguished Career Award. She is recipient of multiple IEEE Best Paper Awards and of the 2012 IEICE Best Paper Award; and recipient of NSERC Discovery Accelerator Supplement Award. She was a Distinguished Lecturer of the IEEE Communications Society (ComSoc) 2013-2016. Prof. Aïssa was a Member-At-Large of ComSoc's Board of Governors 2014-2016. Her editorial activities include: Editorial Board Member, IEEE ACCESS 2022-2024; Editor-At-Large, IEEE TRANSACTIONS ON COMMUNICATIONS 2020-2023; Area Editor, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS 2014-2019; Editor, IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS 2004-2012; Technical Editor, IEEE COMMUNICATIONS MAGAZINE 2004-2015; and Technical Editor, IEEE WIRELESS COMMUNICATIONS MAGAZINE 2006-2010. She has been involved in organizing many flagship conferences of the IEEE, including the 2021 IEEE International Conference on Communications for which she served as the TPC Chair. She is active in promoting women in engineering and is the Founder of the IEEE Women in Engineering Affinity Group in Montreal. She is a Member-at-Large of the ComSoc's Board of Governors 2023-2025.