ICC 2022 - IEEE International Conference on Communications | 978-1-5386-8347-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICC45855.2022.9839204

MCS Selection for Multi-Connectivity and eMBB-URLLC Coexistence in Time-Varying Frequency-Selective Fading Channels

Govindu Saikesava

Neelesh B. Mehta, *Fellow, IEEE*

***Abstract*—Multi-connectivity, in which multiple base stations (BSs) cooperate and jointly transmit to a user, enables a 5G cellular system to meet the challenging reliability requirements of downlink ultra-reliable and low-latency communications (URLLC) traffic. We derive insightful expressions for the achievability, which is the probability that the URLLC user's reliability requirement is met by multi-connectivity. We then propose a low-complexity algorithm to jointly select the set of cooperating BSs and modulation and coding scheme (MCS) to minimize the throughput loss incurred by enhanced mobile broadband (eMBB) users whose time-frequency resources are punctured to carry URLLC data. For time-varying channels with feedback delays, we present a new stochastic reliability requirement for URLLC traffic. The MCS selected on the basis of this requirement markedly increases the probability of meeting the block error rate (BLER) target over the grid of URLLC user locations.**

I. Introduction

5G serves a diverse set of use cases, namely, enhanced mobile broadband (eMBB), massive machine-type communications (mMTC), and ultra-reliable low-latency communications (URLLC). Among these, URLLC presents unique and challenging requirements, which include block error rates (BLERs) as low as 10^{-6} and latencies as small as 1 ms [1]. It enables new applications such as factory automation, telesurgery, and autonomous driving.

New techniques are essential for satisfying URLLC's challenging requirements. One such technique for the downlink is multi-connectivity. In it, multiple base stations (BSs) send the same information to the user to improve the reliability of communication [2]. To achieve a 1 ms latency, the URLLC data spans only two orthogonal frequency division multiplexing (OFDM) symbols. Upon arrival of the URLLC data, the BS immediately transmits it by puncturing or superposing a part of the eMBB data. It also transmits a preemption indicator on the control channel to inform the eMBB user about the time-frequency resources, which are called resource elements in 3GPP, that it punctured [3, Ch. 10]. Other techniques include multi-slot transmissions, in which the transmitter sends multiple copies in consecutive slots, and antenna diversity.

G. Saikesava and N. B. Mehta are with the Dept. of Electrical Communication Eng., Indian Institute of Science, Bangalore, India. Emails: govindu.kesava@gmail.com, nbmehta@iisc.ac.in.

This research was partially supported by the MEITY project "Next Generation Wireless Research and Standardization on 5G and Beyond."

Multi-connectivity can be implemented using either orthogonal transmission (OT) or joint transmission (JT). In OT, the BSs transmit on orthogonal resource elements [4]. In JT, the BSs use maximum ratio transmission (MRT) and transmit on the same resource elements using a common modulation and coding scheme (MCS) [5]. The resultant signal-to-noise ratio (SNR) is larger because it is the sum of individual SNRs from the different BSs. JT requires fewer resources than OT since the BSs transmit on the same resource elements.

A. Focus and Contributions

We focus on multi-connectivity for downlink using JT. We address two important system design issues related to its reliability that have received less attention in the literature. First, while multi-connectivity improves the reliability of URLLC, it requires eMBB resources at multiple BSs to be pre-empted and causes eMBB throughput loss. Therefore, the subset of BSs that cooperate and their MCS need to be jointly optimized. Second, choosing the MCS of the URLLC user as soon as its packet arrives at the BS requires the availability of timely channel state information (CSI) at the BSs. Requesting CSI from the URLLC user on an as-needed basis leads to unacceptably large latencies. Thus, the CSI needs to be fed back periodically by the user. However, this causes the CSI to be partially outdated by the time it is used for transmission.

We make the following contributions:

- When the feedback delays are negligible compared to the coherence time of the channel, we propose a low-complexity multi-connectivity MCS selection algorithm (MCMSA) to select the BSs and their MCS to minimize the sum eMBB throughput loss while satisfying the URLLC error requirement. We also derive insightful expressions for the achievability, which is the probability that a feasible solution exists, i.e., at least one subset of cooperating BSs and MCS can meet the error target.
- In time-varying channels with non-negligible feedback delays, the BSs cannot know the instantaneous SNRs at the time of transmission. Thus, multi-connectivity cannot meet an instantaneous BLER target. We propose a novel stochastic reliability constraint to choose the MCS. It mandates that the probability that the instantaneous BLER at the time of transmission (given the fed back CSI) is less than a target BLER should exceed a pre-specified threshold. We derive a tractable expression for

the above probability. This selection of the MCS meets the error target with a much higher probability than the conventional approach that ignores feedback delays.

To the best of our knowledge, our paper is the first to account for the impact of feedback delays and outdated CSI on multi-connectivity for URLLC. Our formulation also accounts for the 5G requirement that the same MCS must be used for all the subbands on which the URLLC packet is transmitted [6]. This requirement is imposed in the standard to limit control channel overhead even though the SNR varies across subbands due to the channel's frequency-selectivity.

Comparison with Literature: Our work differs from the literature in several respects. We do not assume the linear loss model, in which the eMBB throughput loss is assumed to be directly proportional to the number of punctured resource elements [7]–[10]. In [11], the throughput loss is calculated using mean mutual information per bit. Instead, we employ an accurate simulation-driven approach that systematically tabulates the eMBB throughput-loss of each MCS.

Flat fading is assumed in [5], [7], [9], [12]. On the other hand, we consider the more realistic frequency-selective fading model. While eMBB throughput loss is considered in [7], [11], [12], multi-connectivity is not considered. In [13], eMBB-URLLC coexistence is not considered, and small-scale fading is not modeled in [5], [13]. While coordinated multi-point predates multi-connectivity, issues related to eMBB-URLLC co-existence do not arise in it.

B. Outline and Notation

Section II describes the system model. Section III presents the MCS selection algorithm and achievability analysis when feedback delays are negligible. Section IV addresses the case of non-negligible feedback delays. Section V presents our numerical results. Our conclusions follow in Section VI.

The probability of an event A is denoted by $\Pr(A)$. The conditional probability of A given B is denoted by $\Pr(A \mid B)$. Expectation is denoted by $\mathbb{E}[\cdot]$, and expectation conditioned on Y by $\mathbb{E}[\cdot \mid Y]$. The moment generating function (MGF) of a random variable (RV) X is denoted by $\Psi_X(s)$; it equals $\mathbb{E}[\exp(sX)]$. The size of a set $\mathcal{S}$ is denoted by $|\mathcal{S}|$.

II. System Model

We consider a set $\mathcal{B} = \{1, 2, \ldots, K\}$ of K BSs that serve eMBB users and a URLLC user. The URLLC user is served by a subset $\mathcal{S} \subseteq \mathcal{B}$ of cooperating BSs. An eMBB user is served by one of the BSs. A controller, which is connected to all the BSs, determines $\mathcal{S}$ and the MCS $m \in \mathcal{M}$, where $\mathcal{M} = \{1, 2, \ldots, M\}$ is the set of MCSs, they use to transmit the URLLC packet.

In the 5G physical layer, the system bandwidth is divided into subcarriers in the frequency domain. A group of 12 subcarriers constitutes a physical resource block (PRB). A PRB spans 14 OFDM symbols in time for an eMBB user and 2 OFDM symbols for a URLLC user. q adjacent PRBs are grouped into a subband [6]. The transmission time interval depends on the subcarrier bandwidth. For example, for a subcarrier bandwidth of 15 kHz, it is 1 ms for eMBB and 0.14 ms for URLLC.

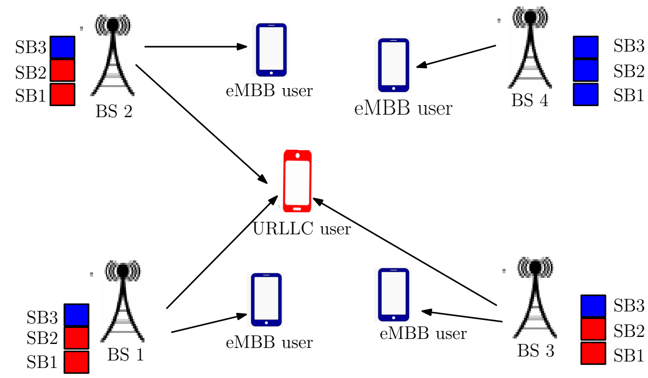


Fig. 1: System model consisting of multiple BSs, that serve a URLLC user and the eMBB users. Red and blue boxes show subbands allocated to the URLLC and eMBB data, respectively.

A. Frequency-Selective Channel Model and Effective SNR

Let H_{ij} denote the channel gain of subband i between the j^{th} BS and the URLLC user. It is a circularly symmetric complex Gaussian RV. Therefore, the channel power gain $|H_{ij}|^2$ is exponentially distributed with mean σ_j^2, which is function of the path-loss between the BS and the user. For tractability, we assume that it is constant over a subband. The fading powers $|H_{ij}|^2$, for $1 \leq i \leq N$ and $1 \leq j \leq K$, are statistically independent. This is justified since the BSs are sufficiently far apart and when the subband bandwidth is comparable to the coherence bandwidth of the channel.

The SNR γ_{ij} for subband i from BS j is given by

$$\gamma_{ij} = \frac{P_T |H_{ij}|^2}{P_T \sum_{k \in \mathcal{B} \setminus \mathcal{S}} |H_{ik}|^2 + \omega^2}, \quad (1)$$

where P_T is the transmit power of a BS per subband, ω^2 is the additive white Gaussian noise (AWGN) power per subband.[1] The BSs in the set $\mathcal{B} \setminus \mathcal{S}$ transmit to eMBB users and thereby cause interference to the URLLC user. In JT, the SNR $\gamma_i^{\mathcal{S}}$ for subband i is equal to [5] $\gamma_i^{\mathcal{S}} = \sum_{j \in \mathcal{S}} \gamma_{ij}$.

In a frequency-selective channel, the URLLC data experiences different SNRs on different subbands. However, as mentioned, the 5G standard requires the same MCS to be used on all the subbands assigned to a user. To systematically evaluate the BLER of any MCS m, we use the effective exponential SNR mapping (EESM), which has been widely used in 3GPP system simulations [14]. For MCS m, EESM maps the vector of SNRs $\gamma_1^{\mathcal{S}}, \ldots, \gamma_{N_m}^{\mathcal{S}}$ to an effective SNR $\zeta_m^{\mathcal{S}}$, where N_m is the number of subbands required for the MCS. $\zeta_m^{\mathcal{S}}$ is the equivalent SNR in a frequency-flat AWGN channel that results in the same BLER. It is given by [15]

$$\zeta_m^{\mathcal{S}} = -\beta_m \log\left(\frac{1}{N_m} \sum_{i=1}^{N_m} \exp\left(-\frac{\gamma_i^{\mathcal{S}}}{\beta_m}\right)\right), \quad (2)$$

[1]For ease of exposition, we do not distinguish between signal-to-interference-plus-noise ratio and SNR.

TABLE I: BLER curve-fit parameters for MCSs specified in 3GPP

MCS	c_m	d_m	ϑ_m (dB)
QPSK, $R = 78/1024$	1.02×10^5	73.22	-8.20
QPSK, $R = 120/1024$	1.97×10^5	67.07	-6.10
QPSK, $R = 173/1024$	7.02×10^5	38.96	-4.61
QPSK, $R = 308/1024$	3.13×10^5	16.27	-1.01
QPSK, $R = 449/1024$	4.97×10^4	9.47	0.70
QPSK, $R = 602/1024$	5.22×10^5	7.42	2.48
16-QAM, $R = 378/1024$	4.50×10^4	3.40	4.97
16-QAM, $R = 490/1024$	4.65×10^4	2.19	6.90
16-QAM, $R = 616/1024$	5.34×10^4	1.46	8.71
64-QAM, $R = 466/1024$	1.56×10^4	0.90	10.30
64-QAM, $R = 567/1024$	8.77×10^3	0.54	12.36
64-QAM, $R = 666/1024$	4.09×10^3	0.29	14.44
64-QAM, $R = 772/1024$	1.86×10^3	0.12	17.94
64-QAM, $R = 873/1024$	91.55	0.04	20.06
64-QAM, $R = 948/1024$	30.10	0.02	21.90

where β_m is an MCS-dependent constant. Its values are tabulated in [16, Tbl. I].

B. BLER Model

BLER is the probability that the data is decoded correctly by the user. Let $\mathsf{BLER}_m(\gamma)$ denote the BLER of MCS m at SNR γ. For the AWGN channel, it is accurately given by the following truncated exponential function [17]:

$$\mathsf{BLER}_m(\gamma) = \begin{cases} 1, & 0 \le \gamma < \vartheta_m, \\ c_m \exp\left(-d_m \gamma\right), & \gamma \ge \vartheta_m, \end{cases} \quad (3)$$

where c_m and d_m are MCS-dependent constants and $\vartheta_m = \log(c_m)/d_m$. Thus, in terms of the effective SNR, the BLER for MCS m is given by $\mathsf{BLER}_m(\zeta_m^{\mathcal{S}})$.

For the different MCSs, which are defined in terms of the modulation scheme and code rate R and are specified in 3GPP in [18, Table 7.2.3.1], Table I shows the curve-fit parameters for a URLLC data payload of 32 bytes, which occupies two OFDM symbols. The number of PRBs assigned depends on the MCS [11]. The parameters are obtained by curve-fitting BLER curves generated from bit-level simulations implemented using Matlab's 5G toolbox. The truncated exponential function approximates the BLER up to three decimal places for all MCSs.

III. With Negligible Feedback Delays

Let L_m denote the eMBB throughput loss when m is the MCS selected for the URLLC user. For an eMBB user, which can be located anywhere in the cell, any of the M MCSs is assumed to be used with equal probability. The eMBB loss depends on the MCS m of the URLLC user because it determines the number of eMBB PRBs that are punctured. The problem of minimizing the eMBB throughput loss can be mathematically stated as follows:

$$\min_{\substack{\mathcal{S} \subseteq \mathcal{B} \\ m \in \mathcal{M}}} \left\{ |\mathcal{S}| L_m \right\}, \quad (4)$$

$$\text{s.t.} \quad \mathsf{BLER}_m\left(\zeta_m^{\mathcal{S}}\right) \le \epsilon, \quad (5)$$

where ϵ is the URLLC error target. Thus, the subset of cooperating BSs $\mathcal{S}$ and the MCS m are jointly optimized.

We first analyze the existence of a feasible solution to this problem. To do this, we analyze *achievability*, which is the probability that at least one subset of cooperating BSs and an MCS can meet the URLLC BLER target.

A. Achievability Analysis

Given the instantaneous channel gains, the URLLC error constraint is satisfied if and only if the BLER when all the BSs transmit with the lowest rate MCS 1 is at most ϵ. This is because the BLER is larger when the set of BSs is smaller or the BSs use a higher rate MCS. The achievability A is given by

$$A = \Pr\left(\mathsf{BLER}_1\left(\zeta_1^{\mathcal{B}}\right) \le \epsilon\right). \quad (6)$$

Substituting the BLER formula in (3), we get $A = \Pr\left(\zeta_1^{\mathcal{B}} \ge \theta\right)$, where $\theta = \log\left(c_1/\epsilon\right)/d_1$.

We now derive an expression for A that captures the effect of various system parameters. Let

$$Y_m = \frac{1}{N_m} \sum_{i=1}^{N_m} \exp\left(-\frac{\gamma_i^{\mathcal{B}}}{\beta_m}\right), \quad (7)$$

denote the term inside the logarithm in (2) with $\mathcal{S} = \mathcal{B}$. Hence,

$$A = \Pr\left(Y_1 \le \exp\left(-\theta \beta_1^{-1}\right)\right). \quad (8)$$

Result *1:* The achievability is given by

$$A = B\left(\exp\left(-\theta \beta_1^{-1}\right), a_1, b_1\right), \quad (9)$$

where $B(.,.,.)$ is the regularized incomplete beta function [19, (6.6.2)]. Here, the beta parameters a_1 and b_1 are given in terms of the two moments of Y_1 by

$$a_1 = \frac{\left(\mathbb{E}\left[Y_1\right]\right)^2 - \mathbb{E}\left[Y_1\right] \mathbb{E}\left[Y_1^2\right]}{\mathbb{E}\left[Y_1^2\right] - \left(\mathbb{E}\left[Y_1\right]\right)^2}, \quad (10)$$

$$b_1 = \frac{\left(1 - \mathbb{E}\left[Y_1\right]\right) a_1}{\mathbb{E}\left[Y_1\right]}. \quad (11)$$

The two moments of Y_1 are, in turn, given by

$$\mathbb{E}\left[Y_1\right] = \prod_{j \in \mathcal{B}} \frac{\lambda_j \beta_1}{\lambda_j \beta_1 + 1}, \quad (12)$$

$$\mathbb{E}\left[Y_1^2\right] = \frac{1}{N_1} \left[\prod_{j \in \mathcal{B}} \frac{\lambda_j \beta_1}{\lambda_j \beta_1 + 2}\right] + \frac{(N_1 - 1)}{N_1} \left(\mathbb{E}\left[Y_1\right]\right)^2, \quad (13)$$

where $\lambda_j = \omega^2/(P_T \sigma_j^2)$.

Proof: The derivation is given in Appendix A. ■

B. Multi-Connectivity MCS Selection Algorithm

In the optimization problem in (4), the number of possible combinations of cooperating BS subsets and MCSs is $(2^{|\mathcal{B}|} - 1)M$. We now propose MCMSA to reduce the computational complexity. Its pseudo-code is given in Algorithm 1. It is based on the following simple idea. For the same SNR, an increase in the MCS index leads to a higher BLER. Hence, if MCS m cannot meet the BLER target, neither can MCS $m+l$,

$\forall l > 0$. Therefore, the BLER for these MCSs need not be computed. We have observed that the number of computations required by MCMSA is lower by a factor of 5 to 7 compared to a brute-force approach depending on P_T, ω^2, M, and K. In case no feasible solution is found, MCMSA selects the lowest rate MCS 1 and the largest subset $\mathcal{S} = \mathcal{B}$.

Algorithm 1 MCMSA

```
Initialization Count = 0.
for every S⊆B do
    for m = 1 : M do
        If BLER_m(ζ_m^S) ≤ ϵ
            • Compute |S|L_m, increase Count by 1.
        Else, if BLER_m(ζ_m^S) > ϵ
            • Move to the next S.
    end for
end for
Select S and m with the smallest sum eMBB throughput loss.
If Count = 0, select the lowest rate MCS 1 for all BSs.
```

IV. With Non-Negligible Feedback Delays

The weights used by the BSs for MRT become partially outdated as they are based on the feedback at time t while transmission occurs at $t+\tau$. We update the notation for the channel gains and SNRs to also show the time indices. We focus on $\mathcal{S} = \mathcal{B}$ since the impact of feedback delays has not been analyzed for this base case.

Let $\mathbf{Q}_t$ denote the CSI fed back at time t. It is a matrix with $(i,j)^{\text{th}}$ element $H_{ij}(t)$. Therefore, the MRT weight of BS j for subband i is $H_{ij}^*(t)/w_i(t)$, where $w_i(t) = \sqrt{\sum_{j\in\mathcal{B}} |H_{ij}(t)|^2}$. Therefore, the SNR $\gamma_i^{\mathcal{B}}(t+\tau)$ of subband i at time $t+\tau$ is

$$\gamma_i^{\mathcal{B}}(t+\tau) = \frac{P_T}{\omega^2 (w_i(t))^2} \left| \sum_{j\in\mathcal{B}} H_{ij}^*(t) H_{ij}(t+\tau) \right|^2. \quad (14)$$

The effective SNR $\zeta_m^{\mathcal{B}}(t+\tau)$ of MCS m at time $t+\tau$ is then

$$\zeta_m^{\mathcal{B}}(t+\tau) = -\beta_m \log\left(\frac{1}{N_m} \sum_{i=1}^{N_m} \exp\left(-\frac{\gamma_i^{\mathcal{B}}(t+\tau)}{\beta_m} \right) \right). \quad (15)$$

.

Since the controller does not know the instantaneous SNRs at time $t+\tau$, it can no longer guarantee the instantaneous BLER target in (5) for the URLLC user. We propose a stochastic reliability constraint for URLLC, as per which the probability that the instantaneous BLER is less than the target value ϵ must be at least $1-\Delta$. The value of $\Delta \ll 1$ depends on the application; a smaller Δ implies a tighter reliability constraint. For MCS m, the constraint can be stated as:

$$\Pr\left(\mathsf{BLER}_m\left(\zeta_m^{\mathcal{B}}(t+\tau)\right) \leq \epsilon \,\middle|\, \mathbf{Q}_t \right) \geq 1-\Delta. \quad (16)$$

Hence, from (3), we get

$$\Pr\left(\zeta_m^{\mathcal{B}}(t+\tau) \geq \theta_m \,\middle|\, \mathbf{Q}_t \right) \geq 1-\Delta, \quad (17)$$

where $\theta_m = \log\left(c_m/\epsilon\right)/d_m$. The challenge lies in computing the conditional probability in (17). Along lines similar to Section III-A, the term inside the logarithm in (15) $Y_m' = \frac{1}{N_m}\sum_{i=1}^{N_m} \exp\left(-\frac{\gamma_i^{\mathcal{B}}(t+\tau)}{\beta_m}\right)$ conditioned on $\mathbf{Q}_t$ can be approximated with a beta RV and yields the following result.

Result *2:* The conditional probability is given by

$$\Pr\left(\zeta_m^{\mathcal{B}}(t+\tau) \geq \theta_m \,\middle|\, \mathbf{Q}_t \right) = B\left(e^{-\frac{\theta_m}{\beta_m}}, a_m', b_m' \right), \quad (18)$$

where the beta parameters a_m' and b_m' are given in terms of the moments of Y_m' conditioned on $\mathbf{Q}_t$ by (10) and (11). And, the conditional moments of Y_m' are given by

$$\mathbb{E}\left[Y_m' \,\middle|\, \mathbf{Q}_t \right] = \frac{1}{N_m} \sum_{i=1}^{N_m} \eta_i^{(1)}, \quad (19)$$

$$\mathbb{E}\left[(Y_m')^2 \,\middle|\, \mathbf{Q}_t \right] = \frac{1}{N_m^2} \left(\sum_{i=1}^{N_m} \eta_i^{(2)} + \sum_{i=1}^{N_m} \sum_{\substack{l=1,\\ l\neq i}}^{N_m} \eta_i^{(1)} \eta_l^{(1)} \right), \quad (20)$$

where $\eta_i^{(1)} = \frac{\beta_m}{\beta_m + 2\alpha_i} \exp\left(\frac{-\delta_i \alpha_i}{\beta_m + 2\alpha_i} \right)$, $\eta_i^{(2)} = \frac{\beta_m}{\beta_m + 4\alpha_i} \exp\left(\frac{-2\delta_i \alpha_i}{\beta_m + 4\alpha_i} \right)$, $\alpha_i = \frac{P_T(1-\rho^2(\tau))}{2\omega^2 w_i^2(t)} \sum_{j\in\mathcal{B}} \sigma_j^2 |H_{ij}(t)|^2$, $\delta_i = \frac{2\rho^2(\tau) w_i^4(t)}{(1-\rho^2(\tau))\sum_{j\in\mathcal{B}} \sigma_j^2 |H_{ij}(t)|^2}$, and $\rho(\tau)$ is the correlation coefficient between $H_{ij}(t)$ and $H_{ij}(t+\tau)$.

Proof: The proof is given in Appendix B. ■

The largest MCS $m \in \mathcal{M}$ that meets the above reliability constraint is selected as it requires the least number of resource elements and causes the smallest eMBB throughput loss. This requires only $O(M)$ computations of the above formula.

V. Numerical Results

We consider a grid of size 300 m $\times$ 300 m. The four BSs are located at the corners. The system bandwidth is 5 MHz, $q = 8$ PRBs, noise figure is 10 dB, noise temperature is 300 K, carrier frequency is 4 GHz, and URLLC data size is 32 bytes [11]. For this payload, it can be verified that only MCSs with rates greater than or equal to QPSK, $R = 308/1024$ can ensure that the URLLC payload fits within two OFDM symbols for the given system bandwidth. There are eleven such MCSs in Table I. For path-loss, we consider the urban macro scenario [6]. We use the Jakes' correlation model. The results are averaged over 500 URLLC user location drops and 2000 fade realizations per each drop.

To determine the eMBB throughput loss for an MCS, we proceed as follows. In a PRB, the resource elements across two OFDM symbols in time and 12 subcarriers in frequency are punctured. The simulations are done using Matlab's 5G toolbox. The increase in BLER is evaluated at the SNR at which the eMBB user's MCS has a BLER of 0.1, which is the BLER target for eMBB in 4G/5G. This approach avoids the inaccuracies in the analytical loss models considered in [7]–[10]. Table II lists the eMBB throughput per PRB without puncturing and the throughput loss per PRB after puncturing for different MCSs. Each PRB is assigned to a different eMBB user. As the MCS rate increases, the eMBB throughput loss increases. For MCSs with rates greater than 64-QAM, $R =$

TABLE II: eMBB throughput and throughput loss per PRB as a function of the MCS used for transmitting eMBB data

MCS	Throughput (kbps)	Throughput loss (kbps)
QPSK, $R = 78/1024$	24	1.03
QPSK, $R = 120/1024$	32	1.25
QPSK, $R = 193/1024$	56	1.62
QPSK, $R = 308/1024$	96	4.51
QPSK, $R = 449/1024$	136	8.02
QPSK, $R = 602/1024$	192	15.74
16-QAM, $R = 378/1024$	240	18.24
16-QAM, $R = 490/1024$	320	26.88
16-QAM, $R = 616/1024$	408	40.80
64-QAM, $R = 466/1024$	456	42.00
64-QAM, $R = 567/1024$	552	60.72
64-QAM, $R = 666/1024$	640	83.20

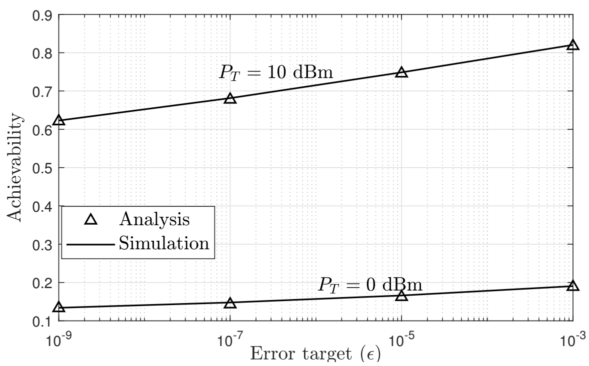


Fig. 2: Achievability as a function of URLLC error target ϵ.

$666/1024$, the BLER is close to 1. These are not shown to conserve space.

A. With Negligible Feedback Delays

Fig. 2 plots the achievability averaged over different URLLC user locations as a function of the error target ϵ for two transmit powers. The achievability increases as ϵ increases. This is because the SNR required to meet the error target decreases. The achievability also increases as P_T increases because the SNR improves.

Fig. 3 benchmarks the total eMBB throughput loss of MCSMA and the conventional approach, in which all the

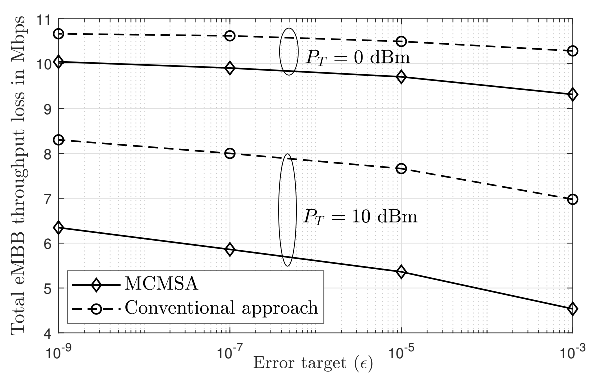


Fig. 3: Total eMBB throughput loss of MCMSA and conventional approach as a function of the target BLER ϵ.

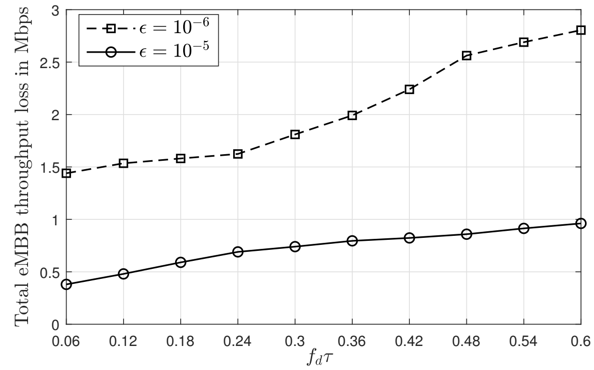
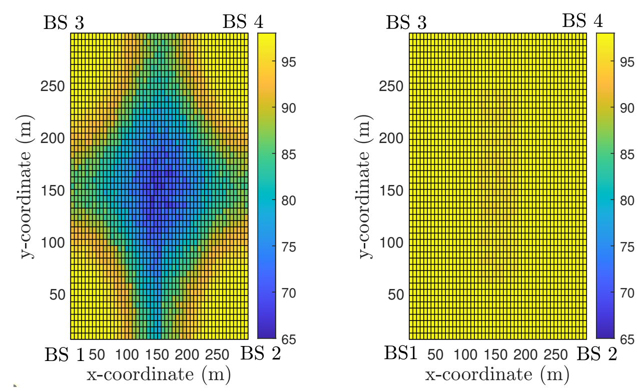


Fig. 4: Total eMBB throughput loss as a function of the normalized feedback delay $f_d\tau$ ($\Delta = 0.05$).



(a) Ignoring delays

(b) Considering delays

Fig. 5: Effect of feedback delays on the probability of meeting error target.

K BSs cooperate. This is done for two values of P_T. As ϵ increases, the throughput loss decreases. This is because the probability that the BSs use a higher rate MCS to transmit the URLLC packet increases, which leads to fewer eMBB resource elements being punctured. MCMSA has a markedly lower throughput loss than the conventional approach.

B. With Non-Negligible Feedback Delays

Fig. 4 plots the total eMBB throughput loss as a function of $f_d\tau$, where f_d is the Doppler spread, for two error targets with $\Delta = 0.05$. The throughput loss increases as $f_d\tau$ increases. This is because a lower rate MCS is selected in order to compensate for the fed back CSI becoming more outdated.

Fig. 5a shows a heat map representing the probability that the instantaneous BLER is below ϵ, at different URLLC user locations when the effect of feedback delays is ignored in the selection of the MCS. At the corners of the grid, the user is close to one of the BSs. Thus, the path-loss from the closest BS is sufficiently small to ensure that the BLER is below ϵ with high probability. However, closer to the grid's center, the path-loss increases and the SNR is more sensitive to multi-path fading. This increases the odds that the BLER exceeds ϵ. Fig. 5b shows the corresponding heat map when the MCS is selected as per the approach proposed in Section IV, which accounts for the feedback delays. Now, the probability that the instantaneous BLER is less than ϵ is markedly higher; it is at least 93% throughout the grid for $\Delta = 0.05$.

VI. Conclusions

For downlink multi-connectivity based on JT, we characterized the probability that the URLLC error requirement could be met given the user location. We proposed a low-complexity algorithm MCMSA to jointly select the set of cooperating BSs and their MCS to minimize the eMBB throughput loss while meeting the URLLC error rate requirements. With non-negligible feedback delays, we saw that the instantaneous BLER requirement could not be satisfied. We proposed a novel stochastic reliability constraint and derived expressions for the conditional probability that the instantaneous BLER was below the target given the fed back CSI. This enabled selection of the MCS increased the probability of meeting the BLER target across the grid of URLLC user locations.

Appendix

A. Proof of Result 1

From (7), Y_1 is a sum of N_1 positive RVs with a finite support of $[0,1]$. As per Papoulis' central limit approximation [15], it can be approximated by a beta RV. The beta distribution parameters a_1 and b_1 can written in terms of the first two moments as per (10) and (11) [20, Ch. 25].

In terms of the beta probability distribution, (8) can be recast as $A = \left(\int_0^{\exp(-\theta/\beta_1)} x^{a_1-1}(1-x)^{b_1-1}\,dx\right)/B(a_1,b_1)$, where $B(.,.)$ is the beta function [19, (6.2.1)]. This yields (9).

Next, we derive expressions for the moments of Y_1. Clearly, $\mathbb{E}[Y_1] = \frac{1}{N_1}\sum_{i=1}^{N_1}\Psi_{\sum_{j\in\mathcal{B}}\gamma_{ij}}\left(-\beta_1^{-1}\right)$. Since γ_{ij} and γ_{il} are independent, $\mathbb{E}[Y_1] = \frac{1}{N_1}\sum_{i=1}^{N_1}\left[\prod_{j\in\mathcal{B}}\Psi_{\gamma_{ij}}(-\beta_1^{-1})\right]$. As γ_{ij} is an exponential RV with parameter λ_j, its MGF can be shown to be $\Psi_{\gamma_{ij}}(s) = \frac{\lambda_j}{\lambda_j - s}$, for $\Re\{s\} < \lambda_j$. Here, $\Re\{.\}$ denotes the real part. Hence, $\Psi_{\gamma_{ij}}\left(-\beta_1^{-1}\right) = \lambda_j/(\lambda_j+\beta_1^{-1})$. Similarly,

$$\mathbb{E}\left[Y_1^2\right] = \frac{1}{N_1^2}\sum_{i=1}^{N_1}\prod_{j\in\mathcal{B}}\mathbb{E}\left[e^{-\frac{2\gamma_{ij}}{\beta_1}}\right] + \frac{1}{N_1^2}\sum_{l=1}^{N_1}\sum_{\substack{i=1,\\ i\neq l}}^{N_1}\left(\prod_{j\in\mathcal{B}}\mathbb{E}\left[e^{-\frac{\gamma_{ij}}{\beta_1}}\right]\right)\left(\prod_{j\in\mathcal{B}}\mathbb{E}\left[e^{-\frac{\gamma_{lj}}{\beta_1}}\right]\right). \quad (21)$$

As above, $\mathbb{E}\left[e^{-\frac{2\gamma_{ij}}{\beta_1}}\right] = \lambda_j/(\lambda_j + 2\beta_1^{-1})$. This yields (13).

B. Brief Proof of Result 2

We can show that $\left|\sum_{j\in\mathcal{B}} H_{ij}^*(t)H_{ij}(t+\tau)\right|$ conditioned on $\mathbf{Q}_t$ is a Rician RV with non-centrality parameter $\rho(\tau)w_i^2(t)$ and scale parameter $\sqrt{\frac{1-\rho^2(\tau)}{2}\sum_{j\in\mathcal{B}}\sigma_j^2|H_{ij}(t)|^2}$. Hence, $\gamma_i^{\mathcal{B}}(t+\tau)$ conditioned on $\mathbf{Q}_t$ is a weighted non-central chi-square RV with weight α_i and non-centrality parameter δ_i, expressions for which are given in the result statement.

Recall that $Y_m' = \frac{1}{N_m}\sum_{i=1}^{N_m}\exp\left(-\frac{\gamma_i^{\mathcal{B}}(t+\tau)}{\beta_m}\right)$. As in Appendix A, the moments of Y_m' can be expressed in terms of the MGF of $\gamma_i^{\mathcal{B}}(t+\tau)$ conditioned on $\mathbf{Q}_t$. It is given by $\Psi_{\gamma_i^{\mathcal{B}}(t+\tau)}(s) = \exp\left(\frac{\delta_i\alpha_i s}{1-2\alpha_i s}\right)/(1-2\alpha_i s)$, for $\Re\{s\} < \frac{1}{2\alpha_i}$.

Conditioned on $\mathbf{Q}_t$, the beta parameters of Y_m' can then be written in terms of its moments as per (10) and (11). As in Appendix A, $\Pr\left(\zeta_m^{\mathcal{B}}(t+\tau) \geq \theta_m \,\middle|\, \mathbf{Q}_t\right)$ can then be written in terms of the incomplete beta function to yield (18).

References

[1] P. Popovski, C. Stefanovic, J. Nielsen, E. de Carvalho, M. Angjelichinoski, K. F. Trillingsgaard, and A. Bana, "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[2] A. Wolf, P. Schulz, M. Dörpinghaus, J. C. S. S. Filho, and G. P. Fettweis, "How reliable and capable is multi-connectivity?" *IEEE Trans. Commun.*, vol. 67, no. 2, pp. 1506–1520, Oct. 2019.

[3] E. Dahlman, S. Parkvall, and J. Sköld, *5G NR: The Next Generation Wireless Access Technology*. Academic Press, 2018.

[4] N. H. Mahmood, A. Karimi, G. Berardinelli, K. I. Pedersen, and D. Laselva, "On the resource utilization of multi-connectivity transmission for URLLC services in 5G new radio," in *Proc. IEEE WCNC*, Apr. 2019, pp. 1–6.

[5] T. Hößler, P. Schulz, E. A. Jorswieck, M. Simsek, and G. P. Fettweis, "Stable matching for wireless URLLC in multi-cellular, multi-user systems," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 5228–5241, May 2020.

[6] "Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.912 (v15.2.0), Jul. 2018.

[7] M. Y. Abdelsadek, Y. Gadallah, and M. H. Ahmed, "Resource allocation of URLLC and eMBB mixed traffic in 5G networks: A deep learning approach," in *Proc. IEEE Globecom*, Dec. 2020, pp. 1–6.

[8] H. Yin, L. Zhang, and S. R. Roy, "Multiplexing URLLC traffic within eMBB services in 5G NR: Fair scheduling," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1080–1093, Nov. 2020.

[9] M. Alsenwi, N. H. Tran, M. Bennis, A. K. Bairagi, and C. S. Hong, "eMBB-URLLC resource slicing: A risk-sensitive approach," *IEEE Commun. Lett.*, vol. 23, no. 4, pp. 740–743, Apr. 2019.

[10] J. Y. R. Liu, G. Yu and G. Y. Li, "Resource management for millimeter-wave ultra-reliable and low-latency communications," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 1094–1108, Feb. 2021.

[11] K. I. Pedersen, G. Pocovi, J. Steiner, and S. R. Khosravirad, "Punctured scheduling for critical low latency data on a shared channel with mobile broadband," in *Proc. IEEE VTC (Fall)*, Sep. 2017, pp. 1–6.

[12] A. Anand, G. de Veciana, and S. Shakkottai, "Joint scheduling of URLLC and eMBB traffic in 5G wireless networks," *IEEE Trans. Networking*, vol. 28, no. 2, pp. 477–490, Feb. 2020.

[13] A. H. Mahdi, T. Hößler, N. Franchi, and G. P. Fettweis, "Multi-connectivity for reliable wireless industrial communications: Gains and limitations," in *Proc. IEEE WCNC*, May 2020, pp. 1–7.

[14] S. S. Tsai and A. C. K. Soong, "Effective-SNR mapping for modeling frame error rates in multiple-state channels," *Tech. Rep. 3GPP2-C30-20030429-010*, Apr. 2003.

[15] J. Francis and N. B. Mehta, "EESM-based link adaptation in point-to-point and multi-cell OFDM systems: Modeling and analysis," *IEEE Trans. Wireless Commun.*, vol. 13, no. 1, pp. 407–417, Dec. 2014.

[16] J. Fan, Q. Yin, G. Y. Li, B. Peng, and X. Zhu, "MCS selection for throughput improvement in downlink LTE systems," in *Proc. ICCCN*, Aug. 2011, pp. 1–5.

[17] Q. Liu, S. Zhou, and G. B. Giannakis, "Cross-layer combining of adaptive modulation and coding with truncated ARQ over wireless links," *IEEE Trans. Wireless Commun.*, vol. 3, no. 5, pp. 1746–1755, Oct. 2004.

[18] "Physical layer procedures," 3rd Generation Partnership Project (3GPP), Tech. Rep. 36.213 (v12.3.0), Oct. 2014.

[19] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, 10th ed. New York, USA: Dover, 1972.

[20] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Continuous Univariate Distributions*, 2nd ed. John Wiley Sons, New York, USA, 1995.