

# Can Massive MIMO Support URLLC?

Hangsong Yan  
NYU Wireless  
New York, US  
hy942@nyu.edu

Alexei Ashikhmin  
Bell Labs, Nokia  
New Jersey, US  
alexei.ashikhmin@nokia-bell-labs.com

Hong Yang  
Bell Labs, Nokia  
New Jersey, US  
h.yang@nokia-bell-labs.com

**Abstract**—We investigate the feasibility of using Massive MIMO to support URLLC in both coherence interval based and 3GPP compliant pilot settings. We consider grant-free uplink transmission with MMSE receiver and adopt 3GPP channel models. In the coherence interval based pilot setting, by extensive system level simulations, we find that using a Massive MIMO base station with 128 antennas and MMSE receiver, URLLC requirements can be achieved in Urban Macro (UMa) Non-Line of Sight (NLoS) with orthogonal pilots and Neyman-Pearson detector. However, in the 3GPP compliant pilot setting, even by using the covariance matrix of Physical Resource Block (PRB) subcarriers for active UE detection and channel estimation as well as open-loop power control, we find that URLLC requirements are still challenging to achieve due to the insufficient pilot length and pilot symbol location regulations in a PRB.

**Index Terms**—URLLC, mMIMO, Active UE Detection

## I. INTRODUCTION

ITU defined three main categories for 5G communications – eMBB (enhanced Mobile BroadBand), URLLC (Ultra-Reliable Low-Latency Communication), and mMTC (massive Machine Type Communications) [1]. URLLC is required by Industry 4.0 applications such as robot motion control and process automation remote control. Its use cases can be established in many industrial sectors including manufacturing, healthcare, transportation, new energy exploration, and entertainment. URLLC specifications include ultra-high reliability (e.g., 99.999%, sometimes also 99.9999% [2]) and ultra-low latency (e.g., 1 ms). At the same time, relatively high data throughput is also required (e.g., 250 bytes per 1 ms [2]). To tackle these challenging requirements, the industry standards have proposed to use grant free (GF) access protocol [3] to reduce scheduling latency. However, this results in multi-UE collision thus requires accurate active UE detection.

We consider the feasibility of using Massive MIMO (mMIMO) technology to support URLLC in two pilot settings. In the coherence interval based pilot setting, both long orthogonal pilots and short non-orthogonal Gold sequence are considered. By incorporating active UE detection, LMMSE channel estimation, and MMSE MIMO receiver, extensive simulations are conducted to quantify the uplink performance of a URLLC system supported by a single mMIMO base station (BS) in both i.i.d. Rayleigh fading and 3GPP channel models [4]. For the 3GPP compliant pilot setting, to tackle with the pilot length limit and the channel coefficients variation across subcarriers in a physical resource block (PRB), we incorporate the channel coefficients covariance matrix across

different subcarriers in a PRB into the active UE detector and LMMSE channel estimator. By applying a 3GPP standards compliant open-loop power control, per-UE effective throughput performance is evaluated numerically.

## II. SYSTEM MODEL

We consider a single cell mMIMO URLLC network. One BS with  $M$  antennas is located at the center of the cell. In total  $\tilde{K}$  single antenna UEs are uniformly distributed in the whole cell. We consider GF access in our work. At the beginning of a certain time duration,  $K$  out of  $\tilde{K}$  UEs are authorized to have GF access for URLLC service and each of these  $K$  UEs is assigned a unique pilot. At a given moment of this time duration,  $K_a$  of  $K$  UEs are active, where  $K_a$  is modeled by a Poisson distribution. We consider orthogonal frequency division multiplexing (OFDM) in our system and the coherence interval length is denoted as  $\tau_c$  in terms of OFDM symbols. We consider two pilot settings. In the first setting, pilots occupy  $\tau$  of  $\tau_c$  OFDM symbols for every subcarrier and the channel coefficient of each subcarrier is constant in a coherence interval and independent in different coherence intervals. We call this setting as coherence interval based pilot setting. In the second 3GPP compliant pilot setting, pilot symbols are distributed in a PRB with an example illustrated in Fig. 4 and channel estimation is performed per PRB. Note that Gold sequence is applied in both settings, but the distribution of pilot symbols in the time-frequency resource is different. Also note that we focus on physical layer analysis on mMIMO supported URLLC in this paper, an analysis on balancing queueing and retransmission can be found in [5].

We denote by  $g_{m,k}^n$  the channel coefficient between the  $m$ th antenna of the BS and the  $k$ th UE at the first subcarrier of the  $n$ th subband. It is modeled by

$$g_{m,k}^n = \sqrt{\beta_k} h_{m,k}^n, \quad (1)$$

where  $\beta_k$  is the large-scale fading coefficient of the  $k$ th UE and  $h_{m,k}^n$  is the small-scale fading coefficient which is antenna and frequency dependent. For a given bandwidth (BW), there are  $N$  independent subbands and the BW of each subband depends on the coherence BW.

## III. COHERENCE INTERVAL BASED PILOTS

### A. Received Signal for UE Detection and Channel Estimation

The received signal at the  $m$ th antenna of the BS and the  $n$ th subband is expressed as

$$\mathbf{y}_m^n = \Phi \mathbf{A} \bar{\mathbf{g}}_{[m]}^n + \mathbf{w}_m^n, \quad (2)$$

where  $\Phi = [\phi_1 \phi_2 \dots \phi_K] \in \mathbb{C}^{\tau \times K}$  is the pilot matrix with  $\tau$  denotes the pilot length and  $\|\phi_k\|_2 = 1$  for  $k = 1, 2, \dots, K$ .  $\mathbf{A}$  is the indicator matrix defined as  $\mathbf{A} = \text{diag}\{[a_1, a_2, \dots, a_K]\}$  with  $a_k = 1$  if the  $k$ th UE is active and  $a_k = 0$  if the  $k$ th UE is not active.  $\bar{\mathbf{g}}_{[m]}^n = \sqrt{\tau \rho_p} [g_{m,1}^n, g_{m,2}^n, \dots, g_{m,K}^n]^T$  where  $\rho_p$  is the normalized signal-to-noise ratio (SNR) of each pilot symbol.  $\mathbf{w}_m^n$  is the noise vector with  $\sim i.i.d. \mathcal{CN}(0, 1)$  entries.

### B. Active UE Detection with Orthogonal Pilots

To use orthogonal pilots, we must have  $\tau \geq K$ . We assume  $\tau = K$  to minimize the pilot overhead. Then  $\phi_k^H \phi_{k'} = \delta(k - k')$  for  $k, k' = 1, 2, \dots, K$  and the optimal Neyman-Pearson (NP) [6] detector can be used for active UE detection. We denote  $\mathbf{f}_k$  as the pre-processing operator for the  $k$ th UE detection. Due to the application of orthogonal pilots, the signal used for the  $k$ th UE detection,  $z_{m,k}^n$ , can be obtained by a simple correlation operation on  $\mathbf{y}_m^n$ :

$$z_{m,k}^n = \mathbf{f}_k^H \mathbf{y}_m^n = \sqrt{\tau \rho_p} \sum_{k' \in \mathcal{A}} \phi_k^H \phi_{k'} g_{m,k'}^n + \phi_k^H \mathbf{w}_m^n, \quad (3)$$

where set  $\mathcal{A}$  is the active UE set. Note that if the  $k$ th UE is active,  $z_{m,k}^n = g_{m,k}^n + \phi_k^H \mathbf{w}_m^n \sim \mathcal{CN}(0, \tau \rho_p \beta_k + 1)$  and if the  $k$ th UE is not active,  $z_{m,k}^n = \phi_k^H \mathbf{w}_m^n \sim \mathcal{CN}(0, 1)$ . Define  $\mathbf{z}_k = [\hat{\mathbf{z}}_{1,k}^1, \hat{\mathbf{z}}_{2,k}^1, \dots, \hat{\mathbf{z}}_{M,k}^1, \hat{\mathbf{z}}_{1,k}^2, \dots, \hat{\mathbf{z}}_{m,k}^2, \dots, \hat{\mathbf{z}}_{M,k}^N]^T \in \mathbb{C}^{MN \times 1}$  where  $\hat{\mathbf{z}}_{m,k}^n = [\text{Re}\{z_{m,k}^n\}, \text{Im}\{z_{m,k}^n\}]$ . Based on the *i.i.d.*  $\mathcal{CN}$  assumption of  $\{h_{m,k}^n\}$ , we can obtain the following two hypotheses about whether the  $k$ th UE is active:

$$\begin{aligned} \mathcal{H}_0 (\text{i.e., inactive}) : \mathbf{z}_k &\sim \mathcal{N}(\mathbf{0}, (\sigma_{0,k}^2/2) \mathbf{I}_{2MN}), \\ \mathcal{H}_1 (\text{i.e., active}) : \mathbf{z}_k &\sim \mathcal{N}(\mathbf{0}, (\sigma_{1,k}^2/2) \mathbf{I}_{2MN}), \end{aligned} \quad (4)$$

where  $\sigma_{0,k}^2 = 1$  and  $\sigma_{1,k}^2 = \tau \rho_p \beta_k + 1$ . We define the test statistics as  $T(\mathbf{z}_k) = \sum_{n=1}^N \sum_{m=1}^M |z_{m,k}^n|^2$ . According to (4) and the NP detector procedures given in [6], the  $k$ th UE is identified as active if

$$T(\mathbf{z}_k) > \gamma', \quad \gamma' = Q_{\chi_{2MN}^2}^{-1}(\text{P}_{\text{FA}}) \sigma_{0,k}^2/2, \quad (5)$$

where  $Q_{\chi_{2MN}^2}(\cdot)$  is the complementary cumulative distribution function of  $\chi_{2MN}^2$ , which is the Chi-square distribution with  $2MN$  degree of freedom and  $\text{P}_{\text{FA}}$  is a predetermined probability of false alarm.

### C. Active UE Detection with Gold Sequence

The application of short non-orthogonal pilots as configured in the 3GPP standards (e.g., Gold sequence) means  $\tau < K$ . Due to this fact, the approach of simple correlation pre-processing + NP detector does not work properly for Gold sequence. In this section, we adopt coordinate-wise descend algorithms from [7] for active UE detection. The detailed form of the algorithm is given in Algorithm 1 where ML, MMV, and NNLS are short for Maximum Likelihood, Multiple Measurement Vector, and Non-Negative Least Squares, respectively. In Algorithm 1,  $\bar{\mathbf{Y}} = \sqrt{\tau \rho_p} \Phi \mathbf{A} \bar{\mathbf{G}}^T + \bar{\mathbf{W}}$  where  $\bar{\mathbf{G}}^T = [(\mathbf{G}^1)^T, (\mathbf{G}^2)^T, \dots, (\mathbf{G}^N)^T]$  and  $(\mathbf{G}^n)^T = [\bar{\mathbf{g}}_{[1]}^n, \bar{\mathbf{g}}_{[2]}^n, \dots, \bar{\mathbf{g}}_{[M]}^n] \in$

$\mathbb{C}^{K \times M}$  is the channel coefficient matrix at the first subcarrier of the  $n$ th subband<sup>1</sup>.

### Algorithm 1 Activity Detection via Coordinate-wise Descend

- 1: **Input:** The empirical covariance matrix  $\hat{\Sigma}_y = \frac{1}{M} \bar{\mathbf{Y}} \bar{\mathbf{Y}}^H$  of the  $\tau \times MN$  matrix of samples  $\bar{\mathbf{Y}}$ .
- 2: **Initialize:**  $\Sigma = \mathbf{I}_\tau, \gamma = \mathbf{0}$ .
- 3: **for**  $i = 1, 2, \dots$  **do**
- 4:   Select an index  $k$  corresponding to the  $k$ th component of  $\gamma = [\gamma_1, \gamma_2, \dots, \gamma_L]^T$  (e.g., randomly).
- 5:   **ML:** Set  $d^* = \max \left\{ \frac{\phi_k^H \Sigma^{-1} \hat{\Sigma}_y \Sigma^{-1} \phi_k - \phi_k^H \Sigma^{-1} \phi_k}{(\phi_k^H \Sigma^{-1} \phi_k)^2}, -\gamma_k \right\}$
- 6:   **MMV:** Set  $d^* = \max \left\{ \frac{\sqrt{\phi_k^H \Sigma^{-1} \hat{\Sigma}_y \Sigma^{-1} \phi_k - 1}}{\phi_k^H \Sigma^{-1} \phi_k}, -\gamma_k \right\}$
- 7:   **NNLS:** Set  $d^* = \max \left\{ \frac{\phi_k^H (\hat{\Sigma}_y - \Sigma) \phi_k}{\|\phi_k\|^4}, -\gamma_k \right\}$
- 8:   Update  $\gamma_k \leftarrow \gamma_k + d^*$ .
- 9:   Update  $\Sigma \leftarrow \Sigma + d^* \times (\phi_k \phi_k^H)$ .
- 10: **end for**
- 11: **Output:** The estimated activity pattern  $\gamma$ .

### D. LMMSE Channel Estimation with Orthogonal Pilots

Since each subcarrier has the same pilot length in the coherence interval based pilot setting, without loss of generality we consider one subcarrier in this section so the notation for subband index is omitted. Then the received pilot signals of all active UEs are

$$\mathbf{Y} = [\mathbf{y}_1 \mathbf{y}_2 \dots \mathbf{y}_m] = \sqrt{\tau \rho_p} \Phi \mathbf{A} \mathbf{G}^T + \mathbf{W}, \quad (6)$$

where  $\mathbf{W} \in \mathbb{C}^{\tau \times M}$  are the noise matrix with  $\sim i.i.d. \mathcal{CN}(0, 1)$  entries.

1) *i.i.d. Rayleigh fading:* When small-scale fading coefficients  $h_{m,k} \sim i.i.d. \mathcal{CN}(0, 1)$  with respect to  $m$  and  $k$ ,  $g_{m,k}$  and  $g_{m',k'}$  are independent if  $m \neq m'$  or  $k \neq k'$ . Based on this property, an LMMSE channel estimator for  $\mathbf{G}$  is given as:

$$\mathbf{E}_{\mathcal{B}} = \sqrt{\tau \rho_p} \Phi_{\mathcal{B}} (\mathbf{C}^{-1} + \tau \rho_p \Phi_{\mathcal{B}}^H \Phi_{\mathcal{B}})^{-1}, \quad (7)$$

where  $\mathcal{B}$  is the predicted active UE set.  $\Phi_{\mathcal{B}} = [b_1 \phi_1 \ b_2 \phi_2 \ \dots \ b_K \phi_K]$  where  $b_k = 1$  if  $k \in \mathcal{B}$  and  $b_k = 0$  if  $k \notin \mathcal{B}$ , and  $\mathbf{C} = \text{diag}\{[\beta_1, \beta_2, \dots, \beta_K]\}$ . Using (7), the estimated channel coefficient matrix  $\hat{\mathbf{G}} \in \mathbb{C}^{K \times M}$  is  $\hat{\mathbf{G}}^T = \mathbf{E}_{\mathcal{B}}^H \mathbf{Y}$ .

2) *3GPP Channel Model based Small-Scale Fading:* When small-scale fading coefficients are generated using 3GPP channel models, channel coefficients of different antennas are correlated. Thus, for an optimal linear channel estimation, the covariance matrices  $\mathbf{C}_k \in \mathbb{C}^{M \times M}, \forall k$  defined below need to be taken into account.

$$\mathbf{C}_k \triangleq \text{Cov}(\mathbf{g}_k, \mathbf{g}_k), \quad \mathbf{g}_k = [g_{1,k}, g_{2,k}, \dots, g_{M,k}]^T. \quad (8)$$

By using the properties of orthogonal pilots, the LMMSE channel estimator for  $\mathbf{g}_k$  is given as

$$\mathbf{E}_k = \sqrt{\tau \rho_p} \mathbf{C}_k (\tau \rho_p \mathbf{C}_k + \mathbf{I}_M)^{-1}. \quad (9)$$

<sup>1</sup>Note that active UE detection via Algorithm 1 requires iteration. In addition, given a target  $\text{P}_{\text{FA}}$  there is no closed-form expression for the test threshold,  $\gamma'$ , which requires numerical method to determine.

TABLE I  
3GPP CHANNEL MODEL PARAMETERS.

Parameter	Value
Channel Name	3D NR UMa NLoS
Carrier Frequency	4.0 GHz
BW (Bandwidth)	40 MHz
Number of antennas ( $M$ )	128
Array setting	$16 \times 8$ Planar Array
sub-carrier spacing	30 KHz
Adjacent antenna spacing	half-wavelength
Adjacent antenna polarization	cross polarization
Polarization angle	$45^\circ$ or $-45^\circ$

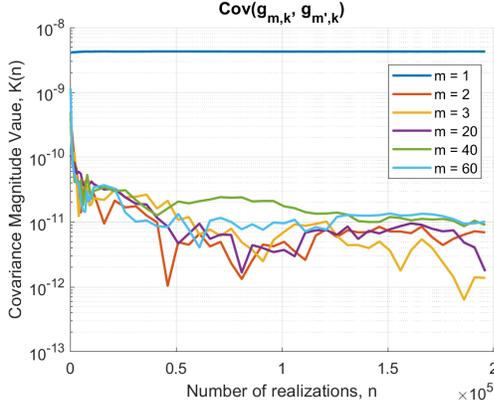


Fig. 1. Channel coefficients covariance value between the first antenna and antennas  $m = 2, 3, 20, 40$ , and  $60$ .

In reality, it might be unpractical to assume that matrices  $\mathbf{C}_k, \forall k$  are available at the BS and hence approximations of  $\mathbf{C}_k, \forall k$  are required. Based on the channel parameter settings given in Table I, we used Nokia 3GPP channel models [4] to generate channel coefficients and found weak channel coefficients correlation among different antennas. Fig. 1 shows an example of the magnitudes of the covariances between one antenna and the other antennas. The  $m = 1$  curve shows the magnitude of the channel coefficient variance. The other curves show the magnitude of channel coefficients covariances between the first antenna and the other antennas. Note that the larger  $m$ , the larger distances between antennas 1 and  $m$ . As we can observe, the magnitudes of the covariances are more than 500 times smaller than the magnitude of the channel coefficient variance. These large gaps indicate low correlations among antennas and thus we approximate  $\mathbf{C}_k$  as

$$\mathbf{C}_k \approx \hat{\mathbf{C}}_k = \text{const}_k \times \mathbf{I}_M, \quad (10)$$

where  $\text{const}_k = \beta_k \text{var}_{3\text{GPP}}$  and  $\text{var}_{3\text{GPP}}$  is the variance of the small-scale fading coefficients generated by the 3GPP channel models. Note that  $\text{const}_k$  will be equal to  $\beta_k$  if small-scale fading coefficients  $\sim i.i.d. \mathcal{CN}(0, 1)$ . The procedures used in 3GPP standards for obtaining  $\beta_k$  such as power head room report can also be used to obtain  $\text{const}_k$ . Thus we assume that the information of  $\text{const}_k$  is available at the BS before the estimation of  $\mathbf{G}$ . Substitute  $\mathbf{C}_k$  with  $\hat{\mathbf{C}}_k$  in (9), we obtain

$\hat{\mathbf{E}}_k = \sqrt{\tau \rho_p} \hat{\mathbf{C}}_k (\tau \rho_p \hat{\mathbf{C}}_k + \mathbf{I}_M)^{-1}$ . Then the estimated channel vector for the  $k$ th UE with  $k \in \mathcal{B}$  is given as:  $\hat{\mathbf{g}}_k = \hat{\mathbf{E}}_k \mathbf{Y}^T \phi_k^*$ . Note that due to the low correlation of the channel coefficients among different antennas, the active UE detection method given by (5) can be directly applied for the 3GPP generated channel coefficients.

#### E. LMMSE Channel Estimation with Gold Sequence

For Gold sequence, we only consider 3GPP channel model for small-scale fading. According to Fig. 1, the correlations among antennas are weak, thus the LMMSE channel estimation approach given in Section III-D1 can be adopted for Gold sequences. Specifically, when using 3GPP channel models, matrix  $\mathbf{C}$  in (7) should be changed to  $\mathbf{C} = \text{diag}\{\{\text{const}_1, \text{const}_2, \dots, \text{const}_K\}\}$ .

### IV. THROUGHPUT WITH COHERENCE INTERVAL BASED PILOTS

#### A. Throughput Computation

We adopt optimal MMSE MIMO receiver in our design and denote by  $\mathbf{v}_k^{\text{MMSE}}$  the receiver vector for the  $k$ th UE. Due to the ultra-low latency requirement of URLLC, instantaneous SINR is adopted for performance evaluation and the expression for the  $k$ th UE is given by:

$$\text{SINR}_k^{\text{inst}}(\boldsymbol{\eta}) = \frac{\rho_u \eta_k (\mathbf{v}_k^{\text{MMSE}})^H \mathbf{g}_k \mathbf{g}_k^H \mathbf{v}_k^{\text{MMSE}}}{(\mathbf{v}_k^{\text{MMSE}})^H \left( \rho_u \sum_{k' \neq k, k' \in \mathcal{A}} \eta_{k'} \mathbf{g}_{k'} \mathbf{g}_{k'}^H + \mathbf{I}_M \right) \mathbf{v}_k^{\text{MMSE}}}, \quad (11)$$

where  $\eta_k$  is the power control coefficient for the  $k$ th UE with the constraint  $0 \leq \eta_k \leq 1$ ,  $\rho_u$  is the normalized SNR of each uplink data symbol. Using Rayleigh-Ritz Theorem,  $\mathbf{v}_k^{\text{MMSE}}$  for  $k \in \mathcal{B}$  can be obtained as

$$\mathbf{v}_k^{\text{MMSE}} = \sqrt{\rho_u \eta_k} \left( \rho_u \sum_{k' \in \mathcal{B}} \eta_{k'} \hat{\mathbf{g}}_{k'} \hat{\mathbf{g}}_{k'}^H + \mathbf{I}_M \right)^{-1} \hat{\mathbf{g}}_k. \quad (12)$$

We use effective per-UE throughput as the performance evaluation metric. If the  $k$ th UE belongs to both sets  $\mathcal{A}$  and  $\mathcal{B}$ , then its effective throughput is expressed as

$$\text{Th}_k = \frac{\tau_c - \tau}{\tau_c} \text{BW} \log_2(1 + \text{SINR}_k^{\text{inst}} \times 10^{-1/10}), \quad (13)$$

where  $10^{-1/10}$  accounts for 1dB SINR degradation due to decoding error<sup>3</sup>. If the  $k$ th UE belongs to set  $\mathcal{A}$  and not to set  $\mathcal{B}$ , this UE is mis-detected and its throughput is set as 0.

<sup>2</sup>Based on (11) the perfect MMSE MIMO receiver have the form:  $\sqrt{\rho_u \eta_k} (\rho_u \sum_{k' \in \mathcal{B}} \eta_{k'} \mathbf{g}_{k'} \mathbf{g}_{k'}^H + \mathbf{I}_M)^{-1} \mathbf{g}_k$ . However, both perfect channel state information (CSI) and accurate statistics of the channel estimation error are not available at the BS if small-scale fading coefficients are generated according to 3GPP channel models, so we adopt the form given by (12).

<sup>3</sup>For a channel with a given SINR, we can transmit at any rate  $R = \log_2(1 + \text{SINR}) - \epsilon$  with  $\epsilon$  an arbitrary small positive number, and achieve an arbitrary small probability of decoding error  $P_{de}$  if a good, for example random, infinitely long error correcting code with optimal decoding is used. In reality a finite length code will be applied and we cannot use optimal decoding (complexity is too high). Thus, we assume that we will manage to achieve a target  $P_{de}$  (e.g.,  $10^{-5}$ ) with  $R = \log_2(1 + \text{SINR} \times 10^{-1/10})$ .

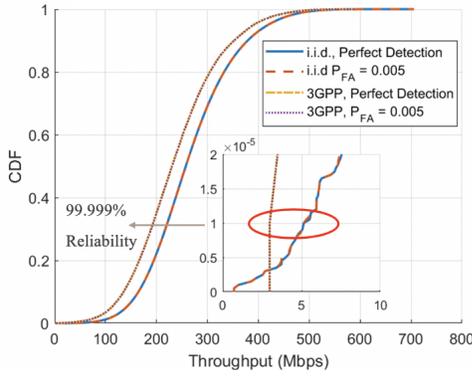


Fig. 2. Per-UE throughput performance with orthogonal pilots (i.e.,  $\tau = 50$ ) under UMa NLoS scenario. Here  $P_{FA}$  is the probability of false alarm.

### B. Numerical Simulation Results

In our simulations the cell radius is set as 150m, the number of URLLC UEs (i.e.,  $K$ ) is set as 50, and at a given moment, the average number of active UEs (i.e.,  $K_a$ ) is set as 10. To enable comparisons with the 3GPP compliant pilot setting presented later, the length  $\tau_c$  of the coherence interval is set as 168 OFDM symbols, which equals the number of resource elements (RE) in a PRB. Note that in practical scenarios,  $\tau_c$  depends on the mobility of environments and can be larger or less than 168. In addition, full power transmission is considered in this section. The per-UE throughput performance with orthogonal pilots under UMa NLoS scenario is shown in Fig. 2. As mentioned in Section I the URLLC requirements are 250 bytes/1ms (2Mbps) with 99.999% reliability. Since 99.999% reliability corresponds to the probability  $10^{-5}$  in a CDF curve, we say that the URLLC requirements are satisfied if the per-UE throughput is larger than 2 Mbps at probability  $10^{-5}$ . As observed from Fig. 2, with orthogonal pilots URLLC requirements are satisfied in both i.i.d. and 3GPP small-scale fading and the probabilities of mis-detection are 0 since no UE has 0 throughput. Furthermore, the performance using NP detector is almost equal to the performance with perfect detection regime, in which we assume that the set of active UEs is known to BS and therefore  $\mathcal{B} = \mathcal{A}$ .

Fig. 3 shows the per-UE throughput performance of the scenario of using Gold sequences under UMa NLoS channels. The pilot length is set as 24 which is the maximum pilot length supported by 3GPP [3] in a PRB, and for comparison the throughput performance of the scenario with orthogonal pilots is also included. It is observed from Fig. 3 that even with perfect detection, the challenging URLLC requirements is not satisfied using Gold sequences. On the other hand, the performance obtained by ML detection is very close to the performance with perfect detection, but MMV and NNLS detection lead to relatively high probability of misdetection ( $P_{MD}$ ) and their performance is farther from the URLLC requirements than the ML detection.

### V. 3GPP COMPLIANT PILOT SETTING

We consider the 3GPP compliant pilot setting in this section. An example of how pilot symbols are allocated in one PRB

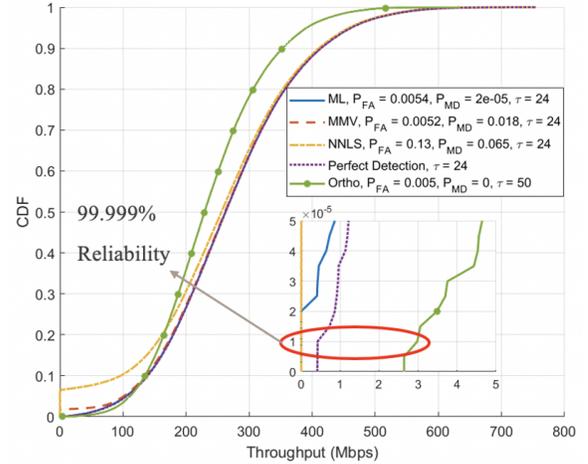


Fig. 3. Per-UE throughput performance with Gold sequences ( $\tau = 24$ ) under UMa NLoS scenario.

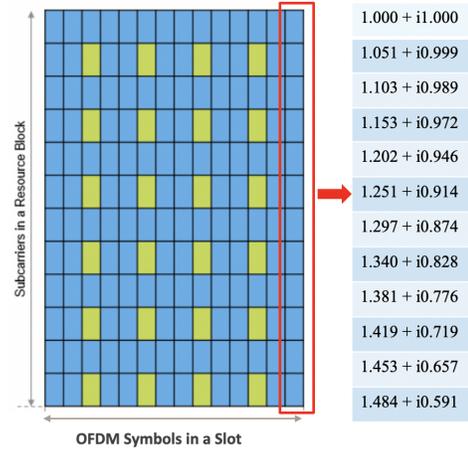


Fig. 4. Pilot symbol locations and normalized channel coefficient values at different subcarriers in one PRB.

is shown in Fig. 4. The normalized channel coefficients at different subcarriers under UMa NLoS channel model are also shown. We assume the channel coefficients stay constant during the time of one slot for every subcarrier and consider the maximum pilot length (i.e., 24). Then the received signal at the  $m$ th antenna in one PRB is given by

$$\mathbf{y}_m = \mathbf{V}_A \tilde{\mathbf{g}}_{[m], \mathcal{A}} + \mathbf{w}_m \quad (14)$$

where the pilot matrix  $\mathbf{V}_A \in \mathbb{C}^{24 \times 6|\mathcal{A}|}$  is given as

$$\mathbf{V}_A = \begin{bmatrix} \mathbf{V}_A^1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{V}_A^3 & \dots & \mathbf{0} \\ \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{V}_A^{11} \end{bmatrix}, \quad (15)$$

where  $\mathbf{V}_A^s = [\mathbf{v}_{\mathcal{A}(1)}^s \mathbf{v}_{\mathcal{A}(2)}^s \dots \mathbf{v}_{\mathcal{A}(|\mathcal{A}|)}^s] \in \mathbb{C}^{4 \times |\mathcal{A}|}$ . Here  $\mathbf{v}_{\mathcal{A}(a)}^s = \phi_{\mathcal{A}(a)}(2(s+1)-3 : 2(s+1)) \in \mathbb{C}^{4 \times 1}$ ,  $s$  denotes the subcarrier index in a PRB and has values  $s = 1, 3, \dots, 11$ ,  $\mathcal{A}(a)$  denotes the  $a$ th component of the active UE set  $\mathcal{A}$  and  $a = 1, 2, \dots, |\mathcal{A}|$  where  $|\mathcal{A}|$  denotes the cardinality of  $\mathcal{A}$ .  $\phi_{\mathcal{A}(a)}(2(s+1) -$

$3 : 2(s+1)$ ) denotes the  $(2(s+1) - 3)$ th to the  $2(s+1)$ th components of the  $\mathcal{A}(a)$ th UE's pilot sequence. The channel vector  $\tilde{\mathbf{g}}_{[m]} = [(\tilde{\mathbf{g}}_{[m],\mathcal{A}}^1)^T (\tilde{\mathbf{g}}_{[m],\mathcal{A}}^3)^T \dots (\tilde{\mathbf{g}}_{[m],\mathcal{A}}^{11})^T]^T \in \mathbb{C}^{6|\mathcal{A}| \times 1}$  with  $\tilde{\mathbf{g}}_{[m],\mathcal{A}}^s = [g_{m,\mathcal{A}(1)}^s, g_{m,\mathcal{A}(2)}^s, \dots, g_{m,\mathcal{A}(|\mathcal{A}|)}^s]^T \in \mathbb{C}^{|\mathcal{A}| \times 1}$ . Here  $g_{m,\mathcal{A}(a)}^s$  denotes the channel coefficient between the  $m$ th antenna of the BS and the  $\mathcal{A}(a)$ th UE at the  $s$ th subcarrier in one PRB.

Since channel estimation is performed per PRB, we can observe from Fig. 4 that at most four pilot symbols can be used for CSI acquisition per subcarrier. Under this condition, a direct application of Algorithm 1 + LMMSE channel estimation in Section III-D1 leads to poor per-UE throughput performance. One possible way to improve the performance is to utilize the channel coefficients covariance matrix across the 12 PRB subcarriers. An estimate of this covariance matrix can be obtained with the help of the sounding reference signals (SRS) defined in the 5G NR standards. Due to space limit, we omit details on obtaining this estimation.

#### A. Active UE Detection

We adopt the ML approach in Algorithm 1 for active UE detection but incorporate the covariance matrix among different subcarriers in a PRB. The procedures are the same as in Algorithm 1 except that step 9 is changed to: Update  $\Sigma \leftarrow \Sigma + \sum_{s=2i+1} \mathbf{C}_k^s \phi_k (\tilde{\phi}_k^s)^H$  for  $i = 0, 1, \dots, 5$ . Here  $\mathbf{C}_k^s \triangleq \text{blkdiag}([c_k^{s,1} \mathbf{I}_4, c_k^{s,3} \mathbf{I}_4, \dots, c_k^{s,11} \mathbf{I}_4]) \in \mathbb{C}^{24 \times 24}$  where  $c_k^{s,s'}$  denotes the covariance between the channel coefficients in the  $s$ th and the  $s'$ th subcarriers in one PRB for the  $k$ th UE with  $s, s' = 1, 3, \dots, 11$ .  $\tilde{\phi}_k^s \in \mathbb{C}^{24 \times 1}$  is a column vector with zero elements except the  $2(s+1) - 3$ th to the  $2(s+1)$ th elements being  $\phi_k(2(s+1) - 3 : 2(s+1))$ .

#### B. LMMSE Channel Estimation

According to the form given by (14), standard LMMSE estimation can be applied to estimate  $\tilde{\mathbf{g}}_{[m]}$  and the corresponding estimator is given as

$$\tilde{\mathbf{E}}_{\mathcal{B}} = \sqrt{\tau \rho_p} \mathbf{V}_{\mathcal{B}} (\tilde{\mathbf{C}}_{\mathcal{B}}^{-1} + \tau \rho_p \mathbf{V}_{\mathcal{B}}^H \mathbf{V}_{\mathcal{B}})^{-1}, \quad (16)$$

where  $\tilde{\mathbf{C}}_{\mathcal{B}} = [(\tilde{\mathbf{C}}_{\mathcal{B}}^1)^T, (\tilde{\mathbf{C}}_{\mathcal{B}}^3)^T, \dots, (\tilde{\mathbf{C}}_{\mathcal{B}}^s)^T, \dots, (\tilde{\mathbf{C}}_{\mathcal{B}}^{11})^T]^T \in \mathbb{C}^{6|\mathcal{B}| \times 6|\mathcal{B}|}$ ,  $\tilde{\mathbf{C}}_{\mathcal{B}}^s = [\mathbf{C}_{\mathcal{B}}^{s,1}, \mathbf{C}_{\mathcal{B}}^{s,3}, \dots, \mathbf{C}_{\mathcal{B}}^{s,s'}, \dots, \mathbf{C}_{\mathcal{B}}^{s,11}] \in \mathbb{C}^{|\mathcal{B}| \times 6|\mathcal{B}|}$ , and  $\mathbf{C}_{\mathcal{B}}^{s,s'} = \text{diag}([c_{\mathcal{B}(1)}^{s,s'}, c_{\mathcal{B}(2)}^{s,s'}, \dots, c_{\mathcal{B}(|\mathcal{B}|)}^{s,s'}])$ . Here  $\mathcal{B}(b)$  denotes the  $b$ th component of the predictive UE set  $\mathcal{B}$  and  $b = 1, 2, \dots, |\mathcal{B}|$ .

#### C. Numerical Simulation Results

In this section, we adopt UMi NLoS channel model with cell radius being 100m and the remaining system settings are the same as in Table I. We also drop 5% UEs with the smallest values of large-scale fading of all  $\tilde{\mathcal{K}}$  UEs. Define by  $|\mathcal{R}|$  the remaining UE set after dropping poor UEs, an open-loop power control is applied which is given as

$$\eta_{\tilde{k}} = \min_{\tilde{k}} \beta_{\tilde{k}} / \beta_{\tilde{k}}, \quad \tilde{k} \in |\mathcal{R}|. \quad (17)$$

Fig. 5 shows the per-UE throughput performance with Gold sequence ( $\tau = 24$ ) under 3GPP compliant and coherence

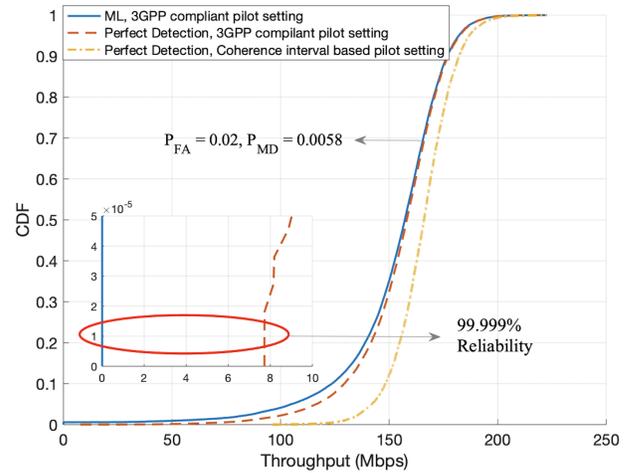


Fig. 5. Per-UE throughput performance with Gold sequence under UMi NLoS scenario. Open-loop power control and dropping 5% poor UEs are applied.

interval based pilot settings. It is observed that the performance with ML detection is very close to that with perfect detection but does not meet the URLLC reliability requirements, which indicates that these requirements can be achieved only for cells of a smaller radius, or number of service antennas  $M$  should be increased. On the other hand, the performance under the coherence interval based pilot setting is significantly higher than the URLLC requirements.

## VI. SUMMARY

We analyzed practical mMIMO supported URLLC in two pilot settings. The simulation results from the coherence interval based pilot setting show that long orthogonal pilots have significant advantage over short non-orthogonal Gold sequences and can meet URLLC requirements even without power control, i.e., with full power transmission, and with relatively large cell radius. In the 3GPP compliant pilot setting, under the assumptions that power control and the subcarrier covariance matrices are used for active UE detection and LMMSE channel estimation, meeting URLLC requirements still appears to be challenging. A reduced cell radius and/or better active UE detection algorithms are required.

## REFERENCES

- [1] ITU, "Minimum requirements related to technical performance for IMT-2020 radio interface(s)," ITU, TR M.2410-0, Nov. 2017.
- [2] 3GPP, "Study on physical layer enhancements for NR Ultra-Reliable and Low-Latency case (URLLC)," 3rd Generation Partnership Project (3GPP), TR 38.824, 16.0.0, Mar. 2019.
- [3] —, "NR; Physical layer procedures for data," 3rd Generation Partnership Project (3GPP), TS 38.214 V16.1.0, Apr. 2020.
- [4] —, "Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), TS 38.901 V16.1.0, Jan. 2020.
- [5] X. Du, Y. Sun, N. B. Shroff, and A. Sabharwal, "Balancing Queuing and Retransmission: Latency-Optimal Massive MIMO Design," *IEEE Transactions on Wireless Communications*, vol. 19, no. 4, pp. 2293–2307, 2020.
- [6] S. M. Kay, "Fundamentals of statistical signal processing: volume II detection theory." Prentice Hall PTR, 1993.
- [7] S. Haghghatshoar, P. Jung, and G. Caire, "Improved Scaling Law for Activity Detection in Massive MIMO Systems," in *2018 IEEE International Symposium on Information Theory (ISIT)*, Vail, CO, 2018, pp. 381–385.