Coursework 3: System-level Performance Evaluation of 5G New Radio MIMO Downlink

Zhaolin Wang

I. Task 1: System Model

This coursework aims at investigating the performance of LTE 4G Single-User MIMO. The base station (BS) servers K user equipments (UEs) in the center cell. The transmission is subject to the interference from the neighboring BS. We assume that there are n_t antennas at the BS and n_r antennas at the UEs.

In the center cell, the BS schedules one UE at a time, and the Spatial Multiplexing with quantized precoding is exploited. The signal transmitted from BS i to user q at time k is given as

$$\mathbf{x}_{q,i} = \underbrace{\mathbf{W}_{q,i}\mathbf{S}_{q,i}^{1/2}}_{\mathbf{P}_{q,i}}\mathbf{c}_{q,i} \tag{1}$$

where the $\mathbf{W}_{q,i} \in \mathbb{C}^{n_t \times r}$ is the precoder selected from the codebook, $\mathbf{S}_{q,i} \in \mathbb{C}^{r \times r}$ is a diagonal matrix controlling the power allocation, and $\mathbf{c}_{q,i} \in \mathbb{C}^{r \times 1}$ are the transmitted symbols. The r is determined by the Rank Indicator (RI) fed back from the user q, and there must be $r \leq \min\{n_r, n_t\}$. The observation at user q is given as

$$\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{x}_{q,i} + \underbrace{\sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{x}_{q,j}}_{\text{inter-cell interference}} + \mathbf{n}_q \tag{2}$$

where the $\Lambda_{q,i}$ is the large scale fading determined by the path loss and the showing, $\mathbf{H}_{q,i} \in \mathbb{C}^{n_r \times n_t}$ is the Rayleigh fading matrix from BS i to user q, and $\mathbf{n}_q \in \mathbb{C}^{n_r \times 1}$ is the noise vector whose entries are the complex Gaussian distribution $\mathcal{CN}(0, \sigma_n^2)$.

A. MMSE Receiver

Define the interference and noise as

$$\begin{aligned} \tilde{\mathbf{n}}_q &= \sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{x}_{q,j} + \mathbf{n}_q \\ &= \sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{P}_{q,j}\mathbf{c}_{q,j} + \mathbf{n}_q \end{aligned} \tag{3}$$

Then the covariance matrix of noise plus interference is denoted by

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{n}}_q} &= \mathbb{E}[\tilde{\mathbf{n}}_q \tilde{\mathbf{n}}_q^H] \\ &= \sum_{j \neq i} \Lambda_{q,j}^{-1}\mathbf{H}_{q,j}\mathbf{P}_{q,j}\mathbf{P}_{q,j}^H\mathbf{H}_{q,j}^H + \sigma_n^2 \mathbf{I}_{n_r \times n_r} \end{aligned} \tag{4}$$

According to (1)(2)(4), the MMSE combiner is given as

$$\begin{aligned} \mathbf{G}_q^{\text{MMSE}} &= \arg\min_{\mathbf{G}_q} ||\mathbf{G}_q\mathbf{y}_q - \mathbf{c}_{q,i}||^2 \\ &= \Lambda_{q,j}^{-1/2}\mathbf{P}_{q,i}^H\mathbf{H}_{q,i}^H(\Lambda_{q,j}^{-1}\mathbf{H}_{q,i}\mathbf{P}_{q,i}\mathbf{P}_{q,i}^H\mathbf{H}_{q,i}^H + \mathbf{R}_{\tilde{\mathbf{n}}_q})^{-1} \end{aligned} \tag{5}$$

Alternatively, considering the l-th stream in the signal, the (2) can be rewritten as

$$\begin{aligned} \mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}c_{q,i,l} + &\underbrace{\sum_{m \neq l} \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}c_{q,i,m}}_{\text{inter-stream interference}} \\ &+ \underbrace{\sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{x}_{q,j}}_{\text{inter-cell interference}} + \mathbf{n}_q. \end{aligned} \tag{6}$$

Denote the noise plus interference for stream l at user q by

$$\tilde{\mathbf{n}}_{q,l} = \sum_{m \neq l} \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}c_{q,i,m} + \sum_{j \neq i} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{x}_{q,j} + \mathbf{n}_q. \tag{7}$$

Its covariance matrix is given by

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{n}}_{q,l}} &= \mathbb{E}[\tilde{\mathbf{n}}_{q,l}\tilde{\mathbf{n}}_{q,l}^H] \\ &= \sum_{m \neq l} \Lambda_{q,i}^{-1}\mathbf{H}_{q,i}\mathbf{p}_{q,i,m}\mathbf{p}_{q,i,m}^H\mathbf{H}_{q,i}^H \\ &+ \sum_{j \neq i} \Lambda_{q,j}^{-1}\mathbf{H}_{q,j}\mathbf{P}_{q,j}\mathbf{P}_{q,j}^H\mathbf{H}_{q,j}^H + \sigma_n^2\mathbf{I}_{n_r \times n_r} \end{aligned} \tag{8}$$

Therefore, the MMSE combiner for stream l is:

$$\mathbf{g}_{q,l}^{\text{MMSE}} = \Lambda_{q,i}^{-1/2}(\mathbf{H}_{q,i}\mathbf{p}_{q,i})^H\mathbf{R}_{\tilde{\mathbf{n}}_{q,l}}^{-1} \tag{9}$$

The output of MMSE at stream l is given as

$$\mathbf{g}_{q,l}^{\text{MMSE}}\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{g}_{q,l}^{\text{MMSE}}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}c_{q,i,l} + \mathbf{g}_{q,l}^{\text{MMSE}}\tilde{\mathbf{n}}_{q,l} \tag{10}$$

According to (10), the output SINR of MMSE at stream l is

$$\begin{aligned} \rho_{q,l} &= \frac{\Lambda_{q,i}^{-1}|\mathbf{g}_{q,l}^{\text{MMSE}}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l}|^2}{\mathbf{g}_{q,l}^{\text{MMSE}}\mathbf{R}_{\tilde{\mathbf{n}}_{q,l}}(\mathbf{g}_{q,l}^{\text{MMSE}})^H} \\ &= \Lambda_{q,i}^{-1}\mathbf{p}_{q,i,l}^H\mathbf{H}_{q,i}^H\mathbf{R}_{\tilde{\mathbf{n}}_{q,l}}^{-1}\mathbf{H}_{q,i}\mathbf{p}_{q,i,l} \end{aligned} \tag{11}$$

The overall achievable rate for user q is

$$R_q = \sum_{l=1}^{r} \log_2(1 + \rho_{q,l}) \tag{12}$$

B. Spatial Multiplexing with Quantized Precoding

The limited amount of feedback at the transmitter is assumed in this coursework. In this case, the quantized precoding is applied, where the precoder $\mathbf{W}_{q,i}$ is selected from a finite set of precoders. This coursework concentrates on the scenario where n_t=4, therefore the LTE 4Tx codebook is used.

At the user q, a Precoding Matrix Indicator (PMI), a Channel Quality indicator, and a Rank Indicator will be decided by the receiver and transmitted back to the BS. The PMI and RI determine which precoder in the codebook will be applied. In this procedure, the precoder is chosen with rank adaptation and rate maximization.

$$\mathbf{W}^*_{q,j} = \arg\max_{r} \max_{\mathbf{W}\in\mathcal{W}_r} R_q \tag{13}$$

where $\mathcal{W}_r$ is the codebook defined for rank r ($r \leq \min\{n_r, n_t\}$).

C. Proportional Fair Scheduling

In the LTE 4G Single-User MIMO, the BS in the center cell schedules one user at a time. In order to guarantee the fairness among the user, the proportioanl fair metric is relied on, where the weighted sum-rate is maximized. At each time instance k, the user q^* is scheduled based on

$$q^* = \arg\max_{q} \frac{R_q(k)}{\bar{R}_q(k)} \tag{14}$$

where $R_q(k)$ is the instantaneous achievable rate of user q and $\bar{R}_q(k)$ is the long-term average rate, which is updated as

$$\bar{R}_q(k+1) = \begin{cases} (1-1/t_c)\bar{R}_q(k) + 1/t_c R_q(k), & q \text{ scheduled} \\ (1-1/t_c)\bar{R}_q(k), & q \text{ not scheduled} \end{cases} \tag{15}$$

where t_c is the scheduling time scale.

D. MIMO Channel Model

The uncorrelated flat fading MIMO channel is assumed following the first-order Gauss-Markov process

$$\tilde{\mathbf{H}}_{k,q,i} = \epsilon\tilde{\mathbf{H}}_{k-1,q,i} + \sqrt{1-\epsilon^2}\tilde{\mathbf{N}}_{k,q,i} \tag{16}$$

The actual MIMO channel is given as

$$\mathbf{H}_{k,q,i} = \tilde{\mathbf{H}}_{k,q,i}\mathbf{R}^{1/2}_{t,q,i} \tag{17}$$

where $\mathbf{R}_{k,q,i}$ is the transmit correlation matrix, which is given as

$$\mathbf{R}_{t,q,i} = \begin{bmatrix} 1 & t_{q,i} & t^2_{q,i} & t^3_{q,i} \\ t^*_{q,i} & 1 & t_{q,i} & t^2_{q,i} \\ t^{2*}_{q,i} & t^*_{q,i} & 1 & t_{q,i} \\ t^{3*}_{q,i} & t^{2*}_{q,i} & t^*_{q,i} & 1 \end{bmatrix} \tag{18}$$

where $t_{q,i} = 0, \forall i \neq 0$ and $t_{q,0} = te^{j\phi_q}$ where t is the magnitude and ϕ_q is the phase randomly distributed between 0 and 2π.

E. User Location

The UEs are randomly and uniformly dropped at a distance $> 35m$ and $< 250m$ from the BS. Figure 1 shows one realization of the UE locations for $K = 10$.

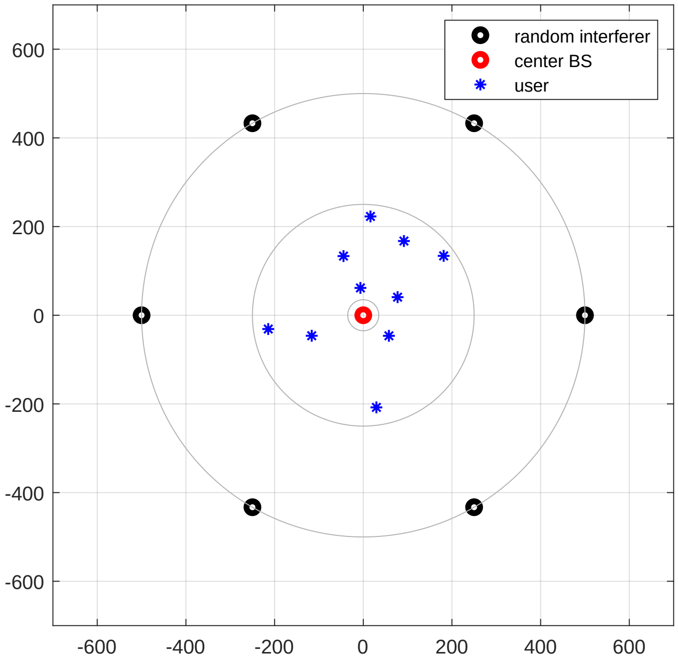

Fig. 1. Locations of UEs ($K = 10$)

II. Task 2: CDF of the long term SINR

For user q in cell i, the long term SINR is given by

$$SINR_{LT,q} = \frac{\Lambda q, i^{-1} E_{s,i}}{\sigma^2_{n,q} + \sum_{j\neq i} \Lambda^{-1}_{q,j} E_{s,j}} \tag{19}$$

In this task, 10^6 user locations are randomly generated and the long term SINR is calculated according to (19).

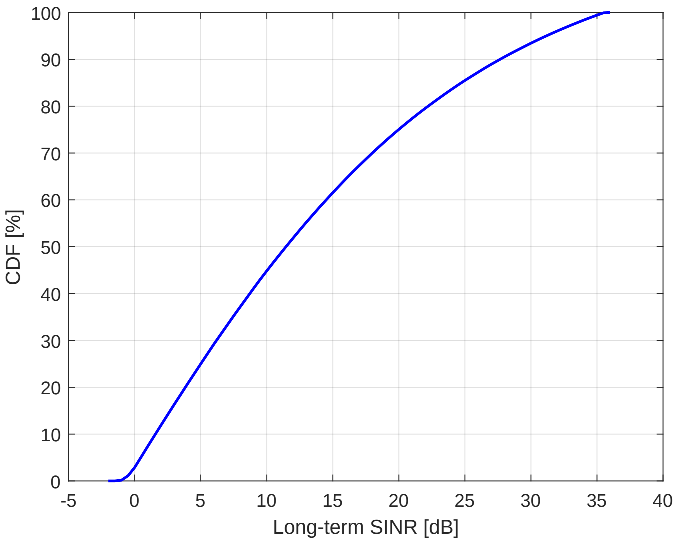

Fig. 2. CDF of the user long term SINR

Figure 2 shows the CDF of the long term SINR. It is clear that the long term SINR is varying from -1.2dB to 35.6dB.

III. Task 3: Influence of the number of receive antennas

For all the following task, the power is uniformly allocated to each streams. The precoders at the neighbouring BS are randomly selected from the codebook. The average user rate is calculated over 800 time instances during which the proportional fairness scheduling is stable. The CDF is calculated by taking 500 random drops of users. The parameters of baseline for all the following tasks are

TABLE I
Baseline Parameters

Parameter	Value
Receive Antennas n_r	1
User Number K	10
Transmit Power P_t	46dBm
Noise Power σ_n^2	-174dBm
Spatial Correlation t	0.5
Time Correlation ϵ	0.85
Scheduling Time Scale t_c	50

Figure 3 shows the CDF of user average rate in the case of different receive antennas. It is obvious that the system have better performance if $n_r = 2$. Compared with $n_r = 1$, the maximum average user rate is increased by 0.7 bps/Hz and minimum average user rate in increased by 0.25 bps/Hz. The reason is that the maximum value of the rank of channel matrix is $r = 2$, therefore the system is capable of supporting 2 streams and there is a multiplexing gain. But for $n_r = 1$, maximum number of streams can be supported is 1, and there is no multiplexing gain exploited. Even though there is only 1 stream transmitted for $n_r = 2$, there is also diversity gain can be exploited at the MMSE receiver, because the MMSE receiver firstly whitens the coloured noise (i.e., interference plus noise) and then performs the maximum ratio combining. Therefore, the case $n_r = 2$ outperforms the case $n_r = 1$ in terms of both multiplexing gain and diversity gain.

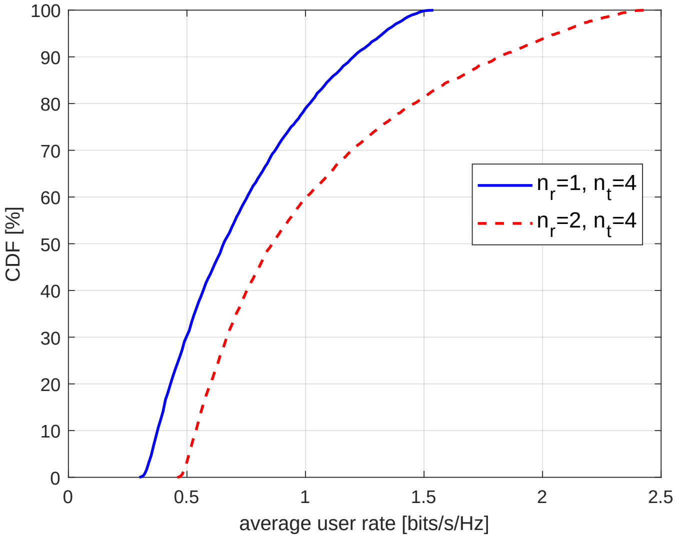

Fig. 3. CDF of user average rate in the case of different receive antennas

IV. Task 4: Influence of the scheduling time scale

Figure 4 shows the CDF of user average rate in the case of scheduling time scales. The average user rate is worst when $t_c = 1.1$, whose minimum value is 0.1 bps/Hz lower than that of $t_c = 50$ and maximum value is 0.2 bps/Hz lower than that of $t_c = 50$. There is also increase of average user rate when t_c increases from 50 to 10^4, but it is not very significant. It is clear that the higher scheduling time scale t_c in (15) results in the better performance of average user rate.

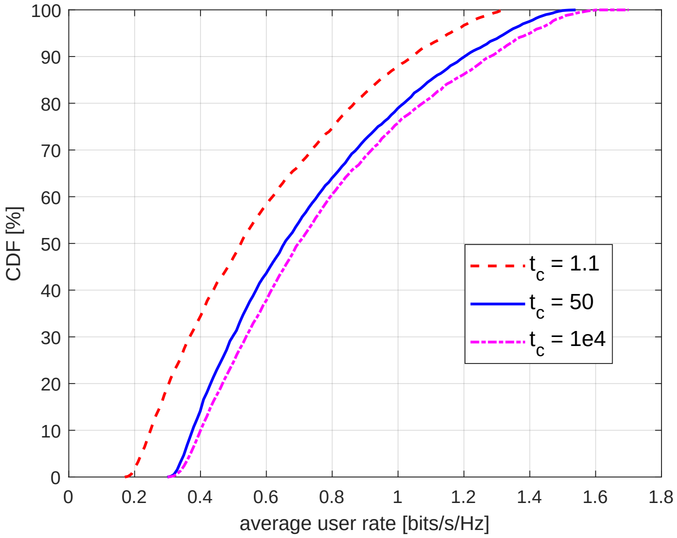

Fig. 4. CDF of user average rate in the case of different scheduling time scales

For the very large t_c, the long-term average rate in (15) becomes

$$\lim_{t_c \to \infty} \bar{R}_q(k+1) = \begin{cases} \bar{R}_q(k), & q \text{ scheduled} \\ \bar{R}_q(k), & q \text{ not scheduled} \end{cases} \tag{20}$$

It means that for the very large t_c, the $\bar{R}_q$ only changes very little at each time instance, and almost keep constant. Therefore, when choosing which user is scheduled at each time instance k based on the criterion (14), the instantaneous rate $R_q(k)$ will dominant the result of selection, i.e., the user with the highest instantaneous rate $R_q(k)$ compared with its $\bar{R}_q$ will be more likely to be scheduled. There will also be multi-user diversity to be exploited. Nevertheless, the drawback of large t_c is the low fairness among the user, as the users with weak channel state have very low opportunity to be scheduled.

For the small t_c, the long-term rate becomes

$$\lim_{t_c \to 1} \bar{R}_q(k+1) = \begin{cases} R_q(k), & q \text{ scheduled} \\ 0, & q \text{ not scheduled} \end{cases} \tag{21}$$

If the user q is not scheduled at time k, the $\bar{R}_q(k+1)$ will directly drop to almost zero. In this case, $R_q(l+1)/\bar{R}_q(k+1) \to \infty$, and the scheduler will equally select from the users which are not scheduled at time k based on (14). Therefore, there is no multi-user diversity gain exploited, resulting in the lower performance of average user rate. However, the fairness is high for the small t_c.

V. Task 5: Influence of velocity/time correlation and the number of users

Figure 5 shows the influence of velocity/time correlation ϵ in (16). One can visualize than with the increase of ϵ, the average user rate is slight decrease. The reason is that the smaller ϵ is, the faster the channel state changes, and vice versa. According to the simulation result, it is obvious that the smaller ϵ leads to the better performance. This is because if the channel states change faster, it is more likely to reach its peak when the user is scheduled, which can lead to the larger achievable rate.

Figure 6 shows the influence of the user number K, where we can conclude that the larger the K is, the lower the performance of average user rate is. From $K = 10$ to $K = 20$, the maximum and minimum average user rate are both halved (from 1.5 bps/Hz to 0.75 bps/Hz and from 0.3 bps/Hz to 0.15 bps/Hz, respectively). The maximum average user rate of $K = 30$ is two thirds of that of $K = 20$, and minimum average user rate also become two thirds. Therefore, the average user rate changes proportionally with K. The reason is that the scheduling time scale is set to $t_c = 50$, which leads to high fairness among all the users. Therefore, the resource allocated to each user is halved as the user number is doubled, resulting in the halved average user rate.

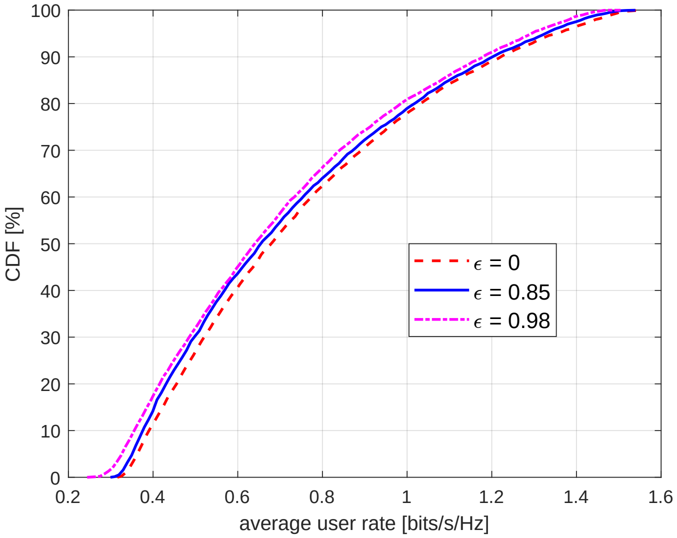

Fig. 5. CDF of user average rate in the case of different velocity/time correlations

VI. Task 6: Influence of the spatial correlation

Figure 7 shows the influence of the spatial correlation t when $n_r = 2$. One can visualize that when t approach one, there is significant decrease of the maximum average user rate, while a slight increase of minimum average user rate. The explanation is given below.

In the case of $n_r = 2$, the channel is a 2×4 MIMO channel and the channel matrix has 2 singular values, which mean there are 2 data streams can be spatially multiplexed in the transmission.

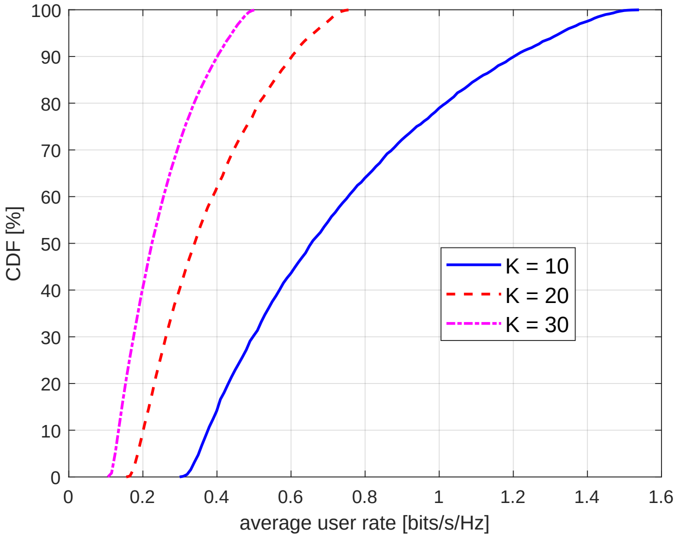

Fig. 6. CDF of user average rate in the case of different user numbers

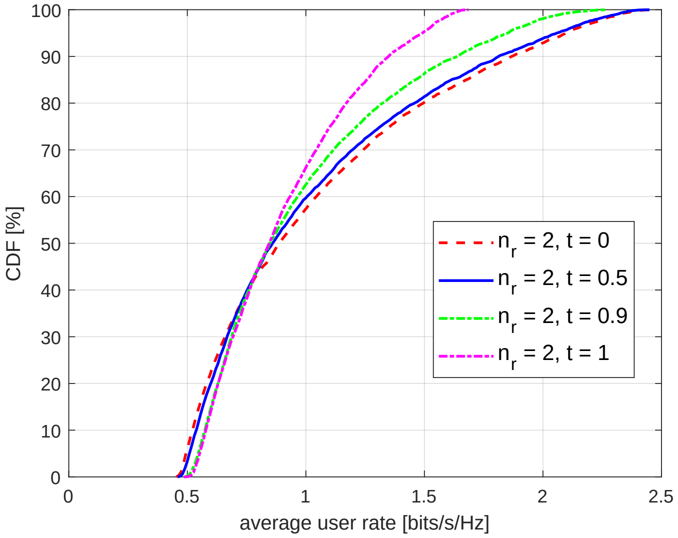

Fig. 7. CDF of user average rate in the case of different spatial correlations for $n_r = 2$

With the increase of the spatial correlation, the rank of the channel matrix approach 1. For the very large spatial correlation t (i.e., $t \to 1$), one of the two singular values is almost zero. We consider the singular value decomposition (SVD) of the 2×4 channel $\mathbf{H}_{q,i}$.

$$\begin{aligned}\mathbf{H}_{q,i} &= \mathbf{U}\begin{bmatrix}\sigma_{max} & 0 & 0 & 0\\ 0 & \sigma_{min} & 0 & 0\end{bmatrix}\mathbf{V}^H\\ &= \mathbf{U}\begin{bmatrix}\sigma_{max} & 0\\ 0 & \sigma_{min}\end{bmatrix}\tilde{\mathbf{V}}^H \qquad (22)\end{aligned}$$

where $\tilde{\mathbf{V}}^H = \mathbf{V}^H(1:2,:)$. Therefore, the observation at (2) can be rewritten as

$$\mathbf{y}_q = \Lambda_{q,i}^{-1/2}\mathbf{U}\begin{bmatrix}\sigma_{max} & 0\\ 0 & \sigma_{min}\end{bmatrix}\tilde{\mathbf{V}}^H\mathbf{x}_{q,i} + \tilde{\mathbf{n}}_q \qquad (23)$$

If we regard $\tilde{\mathbf{V}}^H$ as part of the transmitter and $\mathbf{U}$ as part of the receiver, this 2×4 MIMO channel is converted to 2 virtual SISO channels, the channel gains of which are σ_1^2 and σ_2^2, respectively, i,e.,

$$\tilde{\mathbf{y}}_q = \Lambda_{q,i}^{-1/2} \begin{bmatrix} \sigma_{max} & 0 \\ 0 & \sigma_{min} \end{bmatrix} \tilde{\mathbf{x}}_{q,i} + \tilde{\mathbf{n}}_q \tag{24}$$

where $\mathbf{y}_q = \mathbf{U}\tilde{\mathbf{y}}_q$ and $\tilde{\mathbf{x}}_{q,i} = \tilde{\mathbf{V}}^H \mathbf{x}_{q,i}$. Therefore, there are 2 virtual streams transmitted through the channel no matter the actual stream is 1 or 2. As what is aforementioned, $\sigma_{min} \to 0$ as $t \to 1$. In this case, the channel gain of one of the virtual SISO channel would be very small, and the noise would be always dominant in this virtual SISO channel.

For high SINR and low spatial correlation, σ_{min} is not close to zero, and the uniform power allocation approaches the water-filling solution, which is capacity-achieving power allocation scheme. For the high SINR and high spatial correlation, σ_{min} is almost zero, therefore the virtual SISO channel related to σ_{min} cannot transmit any information. So the maximum average user rate gets smaller when spatial correlation t is larger.

For the low SINR, the channel capacity is mainly determined by σ_{max}. By keeping the channel gain same, the channel with higher spatial correlation will have larger σ_{max}. Therefore, the channel with higher spatial correlation will have larger channel capacity at low SINR. This is why the minimum average user rate of higher spatial correlation t is slightly larger than that of lower t. (One related example is the Task 1 in Coursework 2).

VII. Bonus Task 1: 5G Multi-User Massive MIMO with ZFBF

A. System Model

In this section, we consider the multi-user MIMO BC system with ZFBF. We assume that there are total number of K ($\mathcal{K} = \{1, ..., K\}$) user equipped with single antenna and randomly distributed in the center cell. At each time instance k, the scheduled user set is denoted as $\mathbf{K} \subset \mathcal{K}$. The transmit signal at the BS i is given as

$$\mathbf{x}_i = \sum_{q \in \mathbf{K}} \mathbf{w}_{q,i} s_q^{1/2} c_{q,i} = \mathbf{W}_i \mathbf{S}_i^{1/2} \mathbf{c}_i \tag{25}$$

where $\mathbf{w}_q \in \mathbb{C}^{n_t \times 1}$ is the linear precoder, which will be designed later. s_q and c_q are the power allocate coefficients and information symbols, respectively. $\mathbf{W} = [\mathbf{w}_{p,i}, ...\mathbf{w}_{q,i}]_{p,q\in\mathbf{K}}$, $\mathbf{S} = \text{diag}\left\{s_{p,i}, ..., s_{q,i}\right\}_{p,q\in\mathbf{K}}$, and $\mathbf{c} = [c_{p,i}, ..., c_{q,i}]_{p,q\in\mathbf{K}}^T$.

The observations at all the users are

$$\mathbf{y} = \mathbf{H}_i \mathbf{x}_i + \tilde{\mathbf{n}} = \mathbf{H}_i \mathbf{W}_i \mathbf{S}_i^{1/2} \mathbf{c}_i + \tilde{\mathbf{n}} \tag{26}$$

where

$$\mathbf{y} = [y_p, ..., y_q]_{p,q\in\mathbf{K}}^T \tag{27}$$

$$\mathbf{H}_i = \left[\Lambda_{p,i}^{-1/2} \mathbf{h}_{p,i}^H, ...\Lambda_{q,i}^{-1/2} \mathbf{h}_{q,i}^H\right]_{p,q\in\mathbf{K}}^T \tag{28}$$

$$\tilde{\mathbf{n}} = \sum_{j \neq i} \mathbf{H}_j \mathbf{x}_j + \mathbf{n} \tag{29}$$

The denormalized zero forcing precoder is given as the right pseudo inverse of $\mathbf{H}$

$$\mathbf{F} = \mathbf{H}_i^H (\mathbf{H}_i \mathbf{H}_i^H)^{-1} \tag{30}$$

The normalized precoder $\mathbf{w}_{q,i}$ for user $q \in \mathbf{K}$ is given as

$$\mathbf{w}_{q,i} = \frac{\mathbf{F}(:, q)}{\|\mathbf{F}(:, q)\|} \tag{31}$$

After applying the zero forcing precoder, the intra-cell interference is eliminated at the user. Therefore, the observation at user q is given as

$$\begin{aligned} y_q =& \Lambda_{q,i}^{-1/2} \mathbf{h}_{q,i}^H \mathbf{w}_{q,i} s_{q,i}^{1/2} c_{q,i} \\ &+ \underbrace{\sum_{j \neq i} \Lambda_{q,j}^{-1/2} \mathbf{h}_{q,j}^H \mathbf{W}_j \mathbf{S}_j^{1/2} \mathbf{c}_j}_{\text{inter-cell interference}} + n_q \end{aligned} \tag{32}$$

In this case, the MMSE equalizer at user q is

$$g_q^{\text{MMSE}} = \Lambda_{q,i}^{-1/2} s_{q,i}^{1/2} (\mathbf{h}_{q,i}^H \mathbf{w}_{q,i})^H (R_{\tilde{n}_q})^{-1} \tag{33}$$

where $R_{\tilde{n}_q}$ is the covariance of the noise plus interference, and it is given as

$$R_{\tilde{n}_q} = \sum_{j \neq i} \Lambda_{q,j}^{-1} \mathbf{h}_{q,j}^H \mathbf{W}_j \mathbf{S} \mathbf{W}_j^H \mathbf{h}_{q,j} + \sigma_{n_q}^2 \tag{34}$$

The output SINR of MMSE equalizer is

$$\rho_q = \frac{\Lambda_{q,i}^{-1} s_{q,i} \left|\mathbf{h}_{q,i}^H \mathbf{w}_{q,i}\right|^2}{R_{\tilde{n}_q}} \tag{35}$$

and the achievable rate is

$$R_q = \log(1 + \rho_q) \tag{36}$$

B. Proportional Fair Scheduling for Multi-user

We denote the size of scheduled user set by $|\mathbf{K}|$. In order to make the zero forcing beamforming achievable, there should be $|\mathbf{K}| \leq n_t$. There altogether $C_K^{|\mathbf{K}|}$ possible $|\mathbf{K}|$-subsets of $\mathcal{K}$. Let $\Omega_{|\mathbf{K}|}$ denote the collection of all the possible $|\mathbf{K}|$-subsets. At each time instance k, the BS will select one subset from $\Omega_{|\mathbf{K}|}$. The optimal subset is given as

$$\mathbf{K}^* = \arg \max_{\mathbf{K} \in \Omega_{|\mathbf{K}|}} \sum_{q \in \mathbf{K}} \frac{R_q(k)}{\bar{R}_q(k)} \tag{37}$$

Actually, we also need to choose the best size of the subset $\mathbf{K}$, but in order to simplify the experiment and the reduce the computational cost, the $|\mathbf{K}|$ is fixed in this task.

C. Numerical Results

The parameters in Table I are also used in this simulation, but we assusm that there is no spatial correlation. The size of the subset is set to $|\mathbf{K}| = 4$.

Figure 8 shows the CDF of user average rate in the cases of different number of transmit antennas and users. It is obvious that the performance is degraded with the increase of the user number K, which is the same as the results shown in Figure 6. Besides, the larger number of transmit antennas can lead to

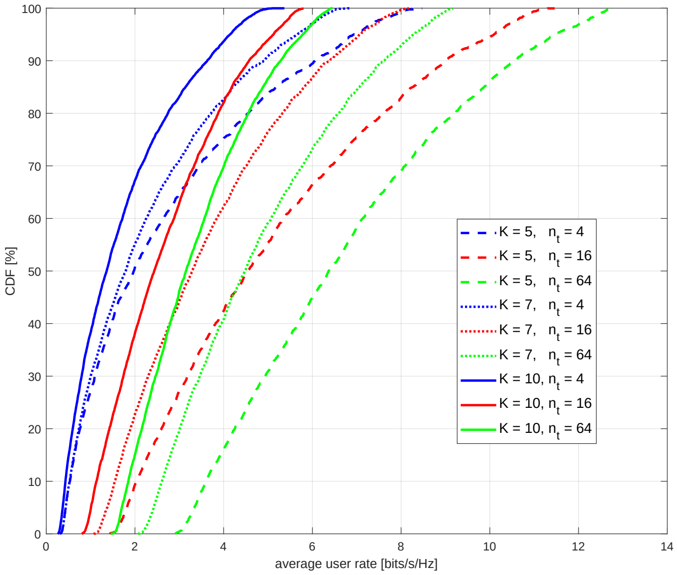

Fig. 8. CDF of user average rate in the cases of different number of transmit antennas and users

the better system performance. This is one of the reasons why we research the massive MIMO.

In massive MIMO, the channel become mutually orthogonal as n_t becomes large, i.e.,

$$\lim_{n_t\to\infty}\frac{1}{n_t}\mathbf{h}_p^H\mathbf{h}_q=\begin{cases}1, p=q\\ 0, p\neq q\end{cases}\tag{38}$$

In this case, the matrix $\mathbf{H}_i$ in (28) will become a unitary-like matrix, i.e.,

$$\mathbf{H}_i\mathbf{H}_i^H=\text{diag}\left\{n_t\Lambda_{p,i}^{-1},...,n_t\Lambda_{q,i}^{-1}\right\}_{p,q\in\mathbf{K}}\tag{39}$$

and the matrix $\mathbf{F}$ in (30) becomes

$$\mathbf{F}=\left[n_t^{-1}\Lambda_{p,i}^{1/2}\mathbf{h}_{p,i},...,n_t^{-1}\Lambda_{q,i}^{1/2}\mathbf{h}_{q,i}\right]\tag{40}$$

Therefore, the zero forcing precoder for user q in $\mathbf{K}$ is

$$\mathbf{w}_{q,i}=\frac{\mathbf{h}_{q,i}}{\|\mathbf{h}_{q,i}\|}\tag{41}$$

where the zero forcing beamforming becomes the matched beamforming.The zero forcing percoder $\mathbf{w}_{q,i}$ is parallel with the channel vector $\mathbf{h}_{q,i}$, due to which the value of $\left|\mathbf{h}_{q,i}^H\mathbf{w}_{q,i}\right|^2$ is maximized. Therefore, the user q has the maximized SINR ρ_q and achievable rate R_q. So as there are more transmit antennas, the channel vector gradually orthogonal to each other, resulting in the better system performance.

VIII. Bonus Task 2: Sum-Rate performance of MISO NOMA and MULP

A. Multi-User MISO NOMA System

Figure 9 shows the K-user MISO NOMA system. The BS is equipped with M antenna and serving K single-antenna users indexed by $\mathcal{K}=\{1,...,K\}$. We assume that $K=gG$ which is grouped in to G groups indexed by $\mathcal{G}=\{1,...,G\}$. The users in the i-th group are indexed by $\mathcal{K}_i=\{ig-g+1,...,ig\}$.

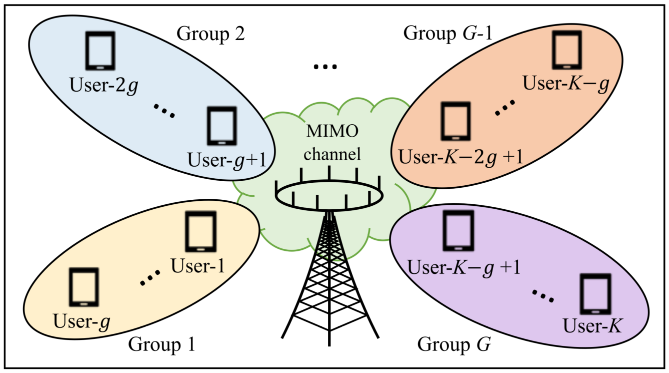

Fig. 9. System architecture with MISO NOMA

The transmit signal is given as

$$\mathbf{x}=\sum_{k\in\mathcal{K}}\mathbf{p}_k s_k\tag{42}$$

where $\mathbf{p}\in\mathbb{C}^{M\times1}$ is the precoder and s_k is the encoded message for user k. The signal receiver at user k is

$$y_k=\mathbf{h}_k^H\mathbf{x}+n_k\tag{43}$$

where $\mathbf{h}_k\in\mathbb{C}^{M\times1}$ is the channel vector and $n_k\sim\mathcal{CN}(0,1)$ is the complex Gaussian noise.

In group $\mathcal{K}_i$, we assume the first user is the strongest user which need to decode the messages of the rest $g-1$ users in this group, the second one is the second strongest user and so on. Therefore, for the user-$j\in\mathcal{K}_i$, it needs to decode the message of user-$\{k|k>j,k\in\mathcal{K}_i\}$. The SINR for user-$j$ to decode the message of user-k is given as

$$\rho_{j,k}=\frac{\left|\mathbf{h}_j^H\mathbf{p}_k\right|^2}{\underbrace{\sum_{m<k,m\in\mathcal{K}_i}\left|\mathbf{h}_j^H\mathbf{p}_m\right|^2}_{\text{intra-group interference}}+\underbrace{\sum_{l\neq i,l\in\mathcal{G}}\sum_{m\in\mathcal{K}_l}\left|\mathbf{h}_j^H\mathbf{p}_k\right|^2}_{\text{inter-group interference}}+1}\tag{44}$$

The related rate is

$$R_{j,k}=\log_2(1+\rho_{j,k})\tag{45}$$

For user-k, in order to make its message decodable at each user which need to decode its messgae, its achievable rate is given as

$$R_k=\min_{j\leq k,j\in\mathcal{K}_i}R_{j,k}\tag{46}$$

If there is no inter-group interference, the $R_{j,k}$ is upper bounded by

$$R_{j,k}\leq\log_2\left(1+\frac{\left|\mathbf{h}_j^H\mathbf{p}_k\right|^2}{1+\sum_{m<k,m\in\mathcal{K}_i}\left|\mathbf{h}_j^H\mathbf{p}_m\right|^2}\right)\tag{47}$$

For R_k, if $R_{ig-g+1,k}$ (the first one in $\mathcal{K}_i$) is the minimum one for all $j\leq k,j\in\mathcal{K}_i$, we have $R_k=R_{ig-g+1,k}$. Otherwise, we have $R_k < R_{ig-g+1,k}$. Therefore, R_k is upper bounded by

$$\begin{aligned} R_k &\le R_{ig-g+1,k} \\ &\le \log_2\left(1+\frac{\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_k\right|^2}{1+\sum_{m=ig-g+1}^{k}\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_m\right|^2}\right) \end{aligned} \tag{48}$$

The sum rate for group $\mathcal{K}_i$ is upper bounded by

$$\begin{aligned} &\sum_{k=ig-g+1}^{ig} R_k \\ &\le \sum_{k=ig-g+1}^{ig}\log_2\left(1+\frac{\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_k\right|^2}{1+\sum_{m=ig-g+1}^{k}\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_m\right|^2}\right) \\ &= \log_2\left(1+\sum_{k=ig-g+1}^{ig}\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_k\right|^2\right) \end{aligned} \tag{49}$$

The R_k can be further upper bounded using Cauchy-Schawarz ineuqality.

$$\begin{aligned} \sum_{k=ig-g+1}^{ig} R_k &\le \log_2\left(1+\sum_{k=ig-g+1}^{ig}\left|\mathbf{h}_{ig-g+1}^H\mathbf{p}_k\right|^2\right) \\ &\le \log_2\left(1+\left\|\mathbf{h}_{ig-g+1}\right\|^2\sum_{k=ig-g+1}^{ig}\left\|\mathbf{p}_k\right\|^2\right) \\ &= \log_2\left(1+\left\|\mathbf{h}_{ig-g+1}\right\|^2 w_i P\right) \end{aligned} \tag{50}$$

where $w_i P$ is the power allocated to the group $\mathcal{K}_i$ and P is the overall transmit power. In this case, the overall sum rate is upper bounded by

$$R_s = \sum_{k=1}^{K} R_k \le \sum_{i=1}^{G}\log_2\left(1+\left\|\mathbf{h}_{ig-g+1}\right\|^2 w_i P\right) \tag{51}$$

Therefore, we can obtain the multiplexing gain of the K-user MISO NOMA system

$$d_s \le \lim_{P\to\infty}\frac{\sum_{i=1}^{G}\log_2\left(1+\left\|\mathbf{h}_{ig-g+1}\right\|^2 w_i P\right)}{\log_2(P)} = G \tag{52}$$

Additionally, for a MIMO system with M transmit antennas and K single-antenna users, the maximum achievable multiplexing gain is $\min\{M,K\}$. So we have $d_s \le \min\{M,K,G\}$. Since $G<K$, the multiplexing gain of the NOMA system is upper bounded by

$$d_s \le \min\{M,G\} \tag{53}$$

B. Multi-User Linear Precoding System

Figure 10 shows the system architecture with MISO MULP. The system has the same setup as the previous MISO NOMA system, but the linear precoding instead of superposition coding is applied at the BS. The transmit signal is given as

$$\mathbf{x} = \sum_{k\in\mathcal{K}}\mathbf{p}_k s_k \tag{54}$$

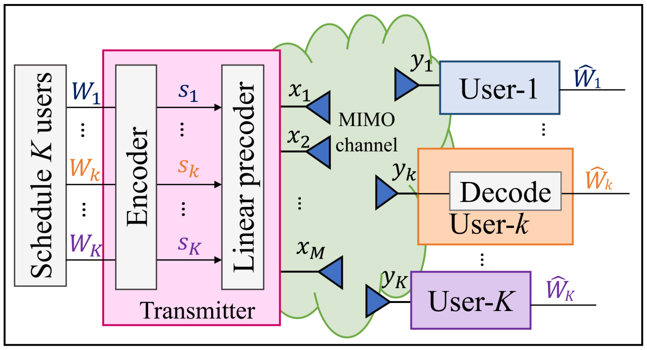

Fig. 10. System architecture with MISO MULP

The signal received at user k is

$$y_k = \mathbf{h}_k^H\mathbf{p}_k s_k + \underbrace{\sum_{j\in\mathcal{K},j\neq k}\mathbf{h}_k^H\mathbf{p}_j s_j}_{\text{interference}} + n_k \tag{55}$$

The SINR at user k is

$$\rho_k = \frac{\left|\mathbf{h}_k^H\mathbf{p}_k\right|^2}{\sum_{j\in\mathcal{K},j\neq k}\left|\mathbf{h}_k^H\mathbf{p}_j\right|^2+1} \tag{56}$$

and the achievable rate is $R_k = \log_2(1+\rho_k)$. We assume the zero forcing precoding is exploited, then we have $\mathbf{h}_k^H\mathbf{p}_j = 0, k\neq j$. The achievable rate can be written as

$$\begin{aligned} R_k &= \log_2(1+\left|\mathbf{h}_k^H\mathbf{p}_k\right|^2) \\ &\le \log_2(1+\left\|\mathbf{h}_k\right\|^2\left\|\mathbf{p}_k\right\|^2) \\ &= \log_2(1+\left\|\mathbf{h}_k\right\|^2 w_k P) \end{aligned} \tag{57}$$

where $w_k P$ is the power allocated to user k and P is the overall transmit power. Therefore, the multiplexing gain of the MULP with ZF is

$$d_s \le \lim_{P\to\infty}\frac{\sum_{k=1}^{K}\log_2(1+\left\|\mathbf{h}_k\right\|^2 w_k P)}{\log_2(P)} = K \tag{58}$$

However, the ZF precoding can support all the K users only if $K\le M$. When $M\le K$, the ZF precoding can only support M users simultaneous and the scheduling is needed, where the multiplexing gain is upper bounded by M. Therefore, we have

$$d_s \le \min\{M,K\} \tag{59}$$

It is obvious that the MULP outperforms the NOMA in terms of multiplexing gain.