System-Level Performance Evaluation of LTE 4Tx MIMO Downlink

Yang Zhao

I. Introduction

Multiple-input multiple-output (MIMO) is widely used in the downlink of contemporary communication systems as long-term evolution (LTE) and Wi-Fi to increase network capacity and resource efficiency. In this article, we investigate the performance of LTE 4Tx single-user MIMO (SU-MIMO) in the broadcast channel (BC) by simulating the cumulative distribution function (CDF) of the user long-term signal-to-interference-and-noise ratio (SINR) and average rate. According to LTE specifications, we use spatial multiplexing (SM) that transmits independent data on different streams and quantised precoding that selects the optimal beamforming vector from the codebook, together with minimum mean squared error (MMSE) receivers, to maximise the instantaneous rate [1]. Fairness is guaranteed by proportional fair (PF) scheduling. It is assumed that there is no delay and error at the feedback, and the fading between users and interference base stations (BS) is not spatially correlated.

II. System Model

A. Deployment and User Distribution

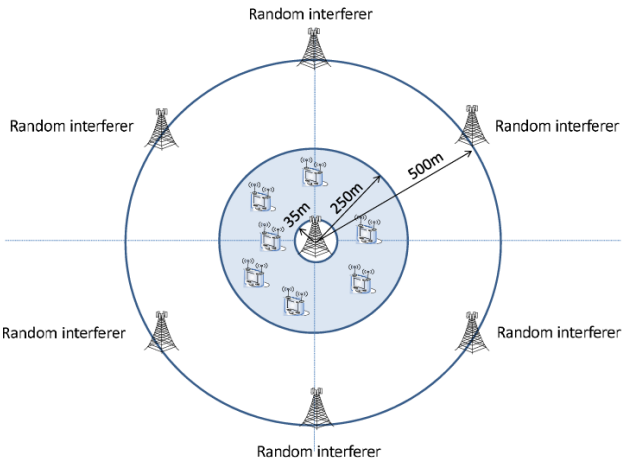

Fig. 1. Deployment scenario of cellular network [2]

As shown in Figure 1, one center BS serves $K = 10$ users randomly and uniformly dropped in an annular region with inner radius $d_{\min} = 35$m and outer radius $d_{\max} = 250$m, surrounded by six interfering cells of distance $d_{bs} = 500$m. All cells are with $n_t = 4$ transmit antennas and users are with $n_r = 1$ or 2 receive antennas. The transmit power is 46dBm at the BSs, while the noise power is -174dBm at the receivers. It is assumed that the drop is long enough to avoid transient state and ensure scheduler converges, but as short as possible so that the movement of users can be neglected in each realisation.

B. Path Loss and Shadowing

Caused by obstacles and blockages, path loss Λ_0 describes the steady decrease in the long-term average of signal power. It can be expressed as a deterministic function of the distance between the transmitter and receiver.

Shadowing S models attenuation by changes in terrain or obstacles. The randomness comes from numerous independent obstructions and can be modelled as a lognormal random variable. It leads to variation in the short-term average of signal power. We treat shadowing as a random process and update it at every time instant.

As major propagation mechanisms, path loss and shadowing results in large-scale signal variations Λ that determines the long-term *(i.e. wideband)* SINR. For user q in cell i, it reads

$$SINR_{LT,q} = \frac{\Lambda_{q,i}^{-1} E_{s,i}}{\sigma_{n,q}^2 + \sum_{j \neq i} \Lambda_{q,j}^{-1} E_{s,j}} \tag{1}$$

C. Correlated Fading

We use a flat fading model with spatial and temporal correlation. The temporally correlated fading $\tilde{\mathbf{H}}_{k,q,i}$ at instant k for user q served by cell i is modelled by a first-order Markov process

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

where $\tilde{\mathbf{H}}_{k-1,q,i}$ is the temporally correlated fading at the previous instant, $\mathbf{N}_{k,q,i}$ is the noise independent to the channel, ε is the time correlation between two instants.

The spatial correlation $\mathbf{R}_{t,q,i}$ of BS i and user q is denoted by a correlation matrix determined by user coordinate

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

where ϕ_q is the angle of user q, $t_{q,0} = te^{j\phi_q}$ for center cell 0 and for simplicity $t_{q,i} = 0$ for interferencing cells $i = 1, \ldots, 6$. Note that the spatial correlation is fixed in each drop. Hence, it is necessary to generate a large number of drops for diversity and convergence. It can be added to the temporally correlated fading by

$$\mathbf{H}_{k,q,i} = \tilde{\mathbf{H}}_{k,q,i} \mathbf{R}_{t,q,i}^{1/2} \tag{4}$$

D. Linear Precoding

For user q of cell i, the diagonal entries of $\mathbf{S}_{q,i}$ denote the power spend on the corresponding layers. Only one user is served at a time in our SU-MIMO system. Assume uniform power allocation and 1 or 2 layers transmission,

$$\mathbf{S}_i = \mathbf{S}_{q,i} = \begin{cases} p_t, & 1-layer \\ \begin{pmatrix} \frac{p_t}{2} & 0 \\ 0 & \frac{p_t}{2} \end{pmatrix}, & 2-layer \end{cases} \tag{5}$$

Also, a transmit beamforming matrix $\mathbf{W}_{q,i}$ is selected from the codebook ω [2] to maximise rate for user q

$$\mathbf{W}_{q,i}^{\star} = \arg\max_{n_e} \max_{\mathbf{W}_{q,i}^{(n_e)} \in \omega_{n_e}} R_q \tag{6}$$

where $n_e = 1, \ldots, \min\{n_t, n_r\}$ is upper-bounded by the rank of channel matrix *(i.e. the number of available streams)*. The outer *max* selects the beamforming matrices to maximise the rate for transmissions with different layers, and their indexes are named precoding matrix indicator (PMI). The inner *max* chooses the number of transmit stream *(i.e. rank indicator, RI)* for rate optimisation. Therefore, the optimum beamforming matrix $\mathbf{W}_{q,i}^{\star}$ guarantees the maximum achievable rate called channel quality indicator (CQI) for user q. For the SU-MIMO system, precoder $\mathbf{P}_{q,i}$ writes as

$$\mathbf{P}_i = \mathbf{P}_{q,i} = \mathbf{W}_{q,i}\mathbf{S}_{q,i}^{1/2} \tag{7}$$

E. MMSE Combiner and Instantaneous Rate

In the SU-MIMO network, a cell serves one user at a time and there is no inter-user *(i.e. intra-cell)* interference. The received signal $\mathbf{y}_{q,l}$ of user $q \in \mathbf{K}_i$ at stream l is

$$\mathbf{y}_{q,l} = \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{P}_{q,i,l}\mathbf{c}_{q,i,l} + \underbrace{\sum_{m \neq l} \Lambda_{q,i}^{-1/2}\mathbf{H}_{q,i}\mathbf{P}_{q,i,m}\mathbf{c}_{q,i,m}}_{interstream} + \underbrace{\sum_{j \neq i}\sum_{l \in \mathbf{K}_j} \Lambda_{q,j}^{-1/2}\mathbf{H}_{q,j}\mathbf{P}_{l,j}\mathbf{c}_{l,j}}_{intercell} + \mathbf{n}_q \tag{8}$$

MMSE combiner is a feasible solution when the noise and interference is colored. To minimise the mean squared error, the optimal weight matrix writes [3]

$$\mathbf{g}_{q,l}^{\star} = \arg\min_{\mathbf{g}_{q,l}} E\left\{\left|\mathbf{g}_{q,l}\mathbf{y}_{q,l} - \mathbf{c}_{q,l}\right|^2\right\} = \Lambda_{q,i}^{-1/2}(\mathbf{H}_{q,i}\mathbf{P}_{q,i,l})^H\mathbf{R}_{n_i}^{-1} \tag{9}$$

where $\mathbf{n}_i$ is noise plus inter-stream and inter-cell interference whose covariance matrix $\mathbf{R}_{\mathbf{n}_i}$ reads

$$\mathbf{R}_{\mathbf{n}_i} = E\left\{\mathbf{n}_i\mathbf{n}_i^H\right\} = \underbrace{\sum_{m \neq l} \Lambda_{q,i}^{-1}\mathbf{H}_{q,i}\mathbf{P}_{q,i,m}(\mathbf{H}_{q,i}\mathbf{P}_{q,i,m})^H}_{interstream} + \underbrace{\sum_{j \neq i} \Lambda_{q,j}^{-1}\mathbf{H}_{q,j}\mathbf{P}_j(\mathbf{H}_{q,j}\mathbf{P}_j)^H}_{intercell} + \sigma_{n,q}^2\mathbf{I}_{n_r,q} \tag{10}$$

The SINR $\rho_{q,l}$ at the output of the MMSE combiner for stream l of user q is

$$\begin{aligned} \rho_{q,l} &= \frac{P_s}{P_{is}+P_{ic}+P_n} = \frac{\Lambda_{q,i}^{-1}|\mathbf{g}_{q,l}\mathbf{H}_{q,i}\mathbf{P}_{q,i,l}|^2}{\mathbf{g}_{q,l}\mathbf{R}_{\mathbf{n}_i}\mathbf{g}_{q,l}^H} \\ &= \Lambda_{q,i}^{-1}(\mathbf{H}_{q,i}\mathbf{P}_{q,i,l})^H\mathbf{R}_{\mathbf{n}_i}^{-1}\mathbf{H}_{q,i}\mathbf{P}_{q,i,l} \end{aligned} \tag{11}$$

Denote the count of employed streams as $n_{u,q}$, the achievable rate $R_{q,i}$ of user q in cell i can be calculated by

$$R_{q,i} = \sum_{l=1}^{n_{u,q}} \log_2(1+\rho_{q,l}) \tag{12}$$

It is also known as instantaneous rate or CQI. MMSE receiver regards all interference as noise, providing a diversity gain of $n_r - n_{u,q} + 1$ that slightly lower than successive interference canceler (SIC) *($n_r - n_{u,q} + i$ for stream i)* and much lower than the optimum maximum likelihood (ML) detector *($n_{u,q}$)* that achieves the ergodic capacity with SM.

F. PF Scheduling and Average Rate

Instead of picking the strongest user, PF scheduler aims to reach a balance between rate and fairness by maximising the weighted sum-rate. The user scheduled at instant k is

$$q^{\star}(k) = \arg\max_q w_q(k)R_q(k) = \arg\max_q \frac{\gamma_q}{\bar{R}_q(k)}R_q(k) \tag{13}$$

Weight $w_q(k)$ depends on the quality of service (QoS) γ_q indicating the priority of user q, together with the long-term average rate

$$\bar{R}_q(k+1) = \begin{cases} (1-1/t_c)\bar{R}_q(k) + 1/t_c R_q(k), & scheduled \\ (1-1/t_c)\bar{R}_q(k), & unscheduled \end{cases} \tag{14}$$

where the scheduling time scale t_c decides the tradeoff between rate and fairness. Notice this long-term average rate pays more attention to the recent performance, while the user average rate is calculated from the instantaneous rate of all moments and drops.

III. Results and Discussion

A. Parameters

Parameters used in the reference model is presented by Table III-A. The number of drops X should have been larger for smoother distributions.

TABLE I
PARAMETERS AS BASELINE

Parameters	Values
Transmit power	$P_t = 46$dBm
Noise variance	$\sigma_n^2 = -174$dBm
Number of user	$K = 10$
Center BS coverage	$d_{\min} = 35$m, $d_{\max} = 250$m
Distance between BSs	$d_{bs} = 500$m
Number of interferencing BS	$n_{interf} = 6$
Antenna configurations	$n_t = 4, n_r = 1$
Path loss model	$\Lambda_0 = 128.1 + 37.6\log_{10}(d/10^3)$dB
Shadowing standard deviation	$\sigma_s = 8$
Time correlation	$\varepsilon = 0.85$
Spatial correlation constant	$t = 0.5$
Drop duration	$T = 10^2$
Scheduling time scale	$t_c = 10^2$
Number of drops	$X = 2 \times 10^2$
QoS	Equal for all users

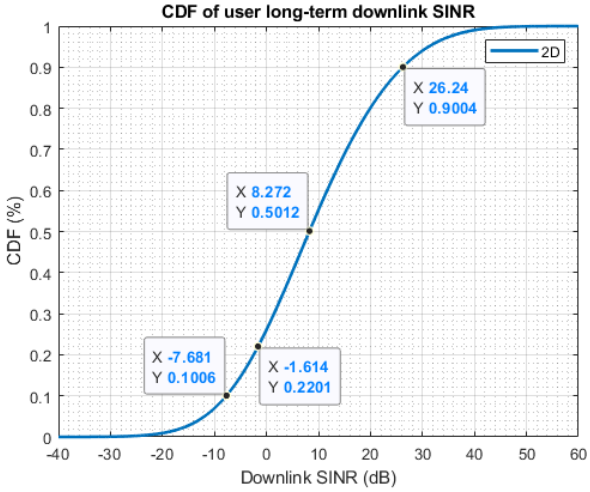

Fig. 2. CDF of user long-term downlink SINR

B. Long-Term SINR

Figure 2 illustrates the impact of large-scale signal variation on the user SINR. If fading is not considered, users with SINR lower than -1.6dB accounts for around 22%, for whom a reliable transmission cannot be established. Also, there are almost a half users with SINR higher than 8.272dB, and a mean value of 8.765dB suggests the SINR distribution is almost symmetrical. Although the coordinate of users are uniformly generated and the distance to the centre cell is with linear CDF, the SINR distribution is a convex-concave function, which brings more users to the middle-SINR region. It can be seen that over 80% users lie in the region where SINR distribution is approximately linear, where the SINR is acceptable for transmission. In other words, the system can guarantee a not low SINR for devices next to the cell edge, and there will be fewer outages. Moreover, it suggests that in this linear region, the SINR is proportional to the distance between the user and the centre cell. Nevertheless, extreme cases may happen with a minor possibility that the SINR may attain 60dB or drop to -20dB when the device is too close to the cell or the edges.

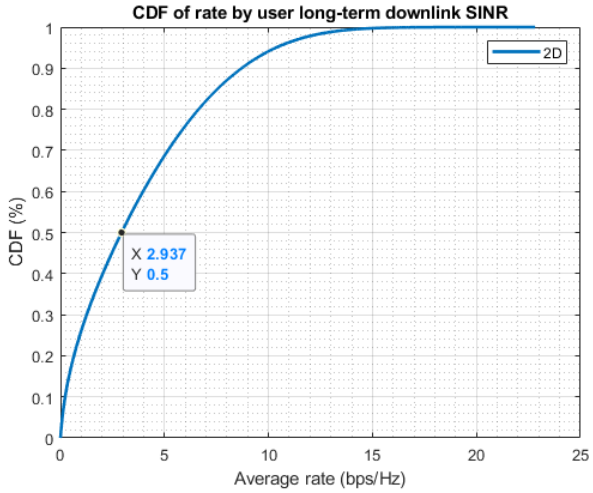

Fig. 3. CDF of instantaneous rate by user long-term downlink SINR

Figure 3 shows the distribution of instantaneous rate by long-term SINR. Without fading, the achievable rate of a half users is lower than 2.937bps/Hz, which is a very small. In the following sections, we will investigate the improvement by fading on instantaneous and user average rate.

C. Number of Receive Antennas

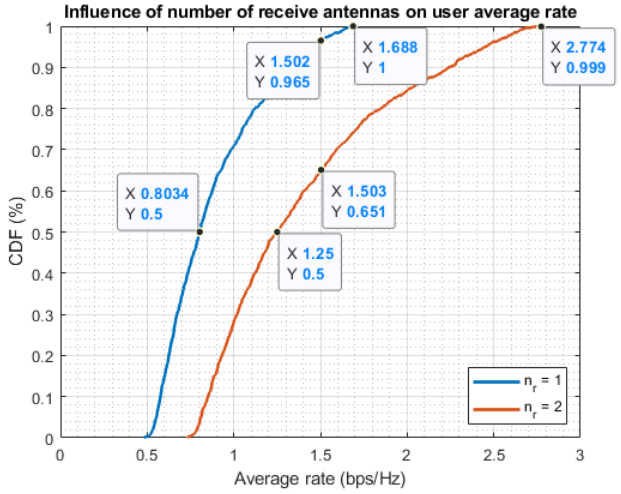

Fig. 4. Influence of the number of receive antennas on user average rate

Figure 4 presents the relationship between receive antennas and average rate. It can be seen that the slope is large in the beginning, meaning more users achieves middle rate and few with low or high rate. For the 1-Rx case, over 50% users achieve an average rate of 0.803bps/Hz. For the 2-Rx case, the value enjoys a 56% increase to 1.25bps/Hz. Also, the percentage of users with a rate larger than 1.5bps/Hz takes up 3.5% with 1-Rx and 35% with 2-Rx. The improvement from an extra antenna is significant.

The 1-Rx case corresponds to the multi-input single-output (MISO) scenario. All power is allocated to the only available stream, and there is no inter-stream interference. With 2 receive antennas, the number of available layers $n_e = \min(n_r, n_t) = 2$. We compare the rates of dominant eigenmode transmission (DET) relying on the stronger layer and multiple eigenmode transmission (MET) using on both streams, then employ the mode that maximises the instantaneous rate. Equation 12 indicates that instead of comparing $\log_2(1+\rho_0)$ and $2\log_2(1+\rho_0/2)$, SINR of each stream is linked to inter-cell interference happening in both modes and inter-stream interference existing in MET. At low SINR, the 2-Rx system tends to transmit one stream which is aligned with the 1-Rx one. In comparison, the rate can be enhanced by using two layers at high SINR because of power saturation.

For the ideal situation of independent streams without all interferences, the cross-over point for DET and MET is 0dB. According to Equation 11, interferences lead to a smaller SINR, and the trade-off becomes more complicated, despite the overall trend of preferring DET at low SINR while favouring MET at high SINR. Moreover, the multiplexing gain can be approximated by the ratio of the maximum rate for which the SINR is sufficiently high, which is $2.774/1.688 = 1.643$. It is smaller than 2 because of the interference and variation of SINR.

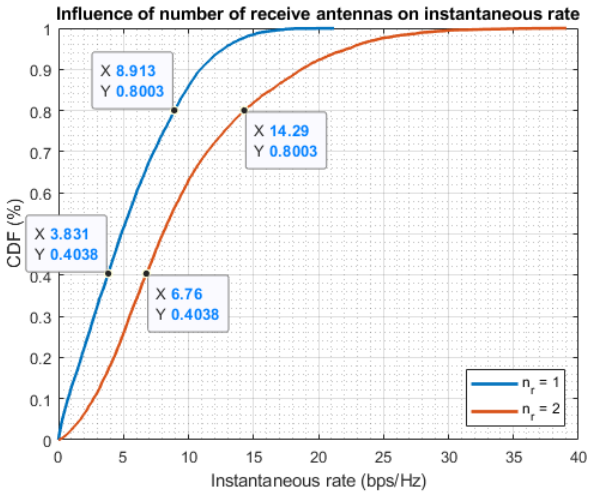

Fig. 5. Influence of the number of receive antennas on instantaneous rate

Distribution of the instantaneous rate is shown in Figure 5. Compared with the no-fading case of Figure 3, it can be concluded that fading brings more diversity and capacity to the channel. Also, the gap between the 1-Rx and 2-Rx scenario is widened as rate and SINR increases. It demonstrates the statement that rate impairment by saturation at high SINR can be mitigated by enabling the second stream.

D. Scheduling Time Scale

Figure 6 presents the impact of scheduling time scale t_c. For $t_c = 1.1$, half users are with average rate lower than 0.5bps/Hz, which equals the lower bound of $t_c = 10$. As mentioned above,

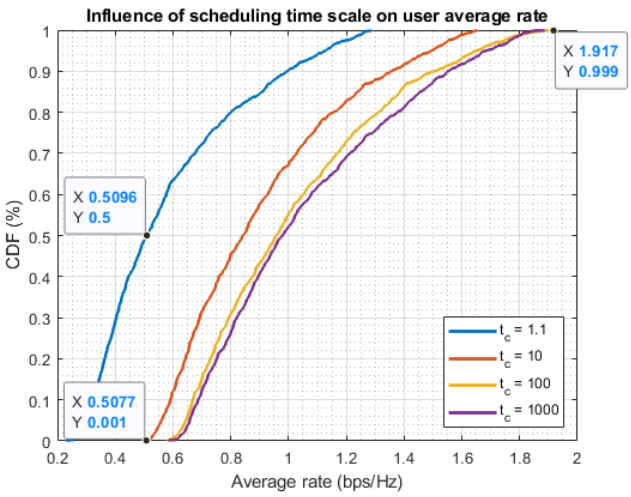

Fig. 6. Influence of the scheduling time scale on user average rate

t_c indicates the "memory" of the scheduler when balancing rate and fairness. If t_c is very small, the scheduler cares more on a short interval in the past, and the long-term average rates of the users served less often will be relatively small. It leads to extremely large weights $w_q = \gamma_q / \bar{R}_q$ that give the resource to user q, although the instantaneous rate may be small. It tends not scheduling the strongest user but to serve those not scheduled in the past. In other words, the importance of weight overpowers the achievable rate, and the PF scheduling with short t_c boils down to *Round-Robin* that serves all users in turn. As denoted by the blue curve, the overall average rate of $t_c = 1.1$ is significantly smaller than others, although it guarantees absolute fairness.

In comparison, the influence of scheduling a particular user on the long-term average rate is negligible if the scheduling time scale is sufficiently large. For $t_c > 100$, the impact of further increasing t_c on average rate is negligible. The reason is that a large denominator makes little difference in the long-term rate, and the stability of the long-term average rate leads to little fluctuations of weights, whose impact on scheduling can be neglected if all users have the same QoS. In other words, the scheduling is dominated by the available rate, and users are scheduled when reaching peaks of the channel. Therefore, PF scheduler with a large t_c essentially serves the strongest user with the largest instantaneous rate, which is equivalent to *rate-maximisation*. Although the rate is improved, the system is unfair, and users may need to wait long before being scheduled. It is recommended to choose a t_c between 10 and 100 to achieve a balance between fairness and performance.

E. Number of Users and Time Correlation

With a fixed resource, serving more users means less average rate. With K users in a cell, the average scheduling duration for each user is $1/K$. On the contrary, the multiuser diversity is expected to provide an SNR gain of $\log K$ for the strongest user when the fading is uncorrelated. Nevertheless,

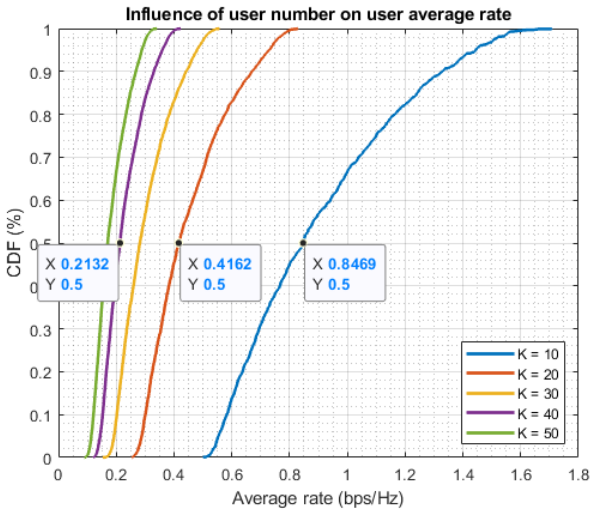

Fig. 7. Influence of the number of users on average rate ($\varepsilon = 0.85$)

Figure 7 demonstrates that with a time correlation of 0.85, the enhancement by multiuser diversity is negligible. As K increases from 10 to 20, the upper bound of the average rate of half users is approximately halved, which is similar to the trend of doubling k from 20 to 40.

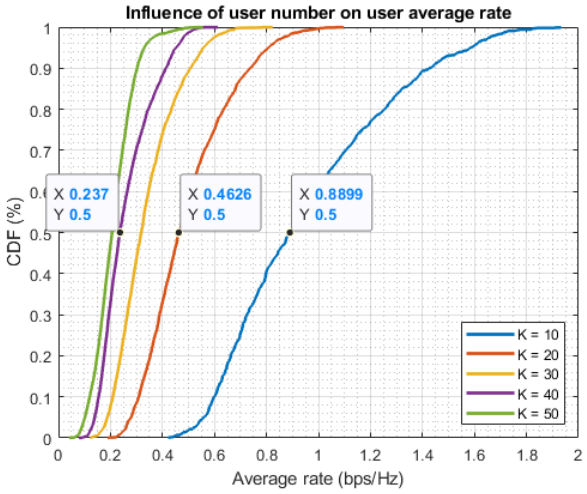

Fig. 8. Influence of the number of users on average rate ($\varepsilon = 0$)

Figure 8 proves that the average rates are improved when there is no correlation in time. The optimisation comes from the enhanced multiuser diversity from temporally uncorrelated fading, which is 5% and 11% for the 10-user and 40-user scenarios. It demonstrates that independent fading can enhance the multiuser diversity to benefit the performance further. However, it is estimated that the advantage of multiuser diversity should decrease for MU-MIMO systems where inter-user interference is considered.

Time correlation ε is a measure of the velocity of channel variation. If $\varepsilon = 1$, fading is static in each drop because the spatial correlation is also fixed. Therefore, there is no

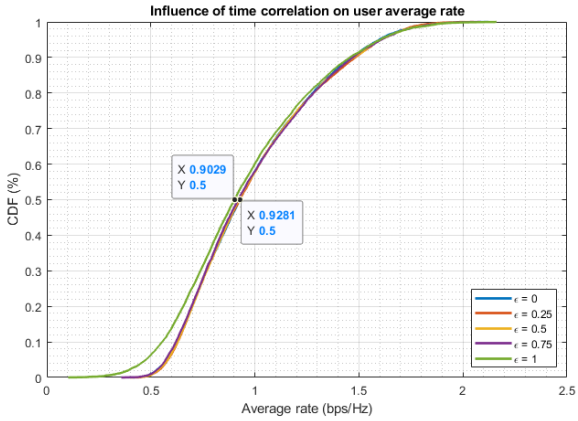

Fig. 9. Influence of the time correlation on user average rate

randomness in fading, but the innovation is still provided by shadowing as a random process. In contrast, the channel is temporally independent when $\varepsilon = 0$ that changes very fast over time, leading to a larger possibility of seeing a higher channel peak in the next time instant. According to Figure 9, the rate distributions are almost the same except the static case that produces the smallest average rate. For the static fading, the chance of having a better channel depends completely on shadowing. On the other hand, as long as there exists alternation in fading, that possibility tends to grow, and a larger rate can be attained. It also suggests that in real-world cases where ε cannot be 1, the actual value plays a minor role in the average rate with a long-enough statistical period. Nevertheless, the instantaneous rate and user experience are related to the time correlation. For a fixed t_c, a larger ε leads to slower channel variations, and the instantaneous rates fluctuate slowly. Hence, strong users are more likely to be scheduled successively, while weak users tend to wait long for peaks. In other words, a large temporal correlation may result in fairness issue that can be solved by balancing ε and t_c.

F. Spatial Correlation

The impact of the spatial correlation on user average rate for 1-Rx and 2-Rx models are shown in Figure 10 and 11. For the 1-Rx scenario, the average rate seems irrelevant to the spatial correlation constant t. In contrast, when there are 2 receive antennas, the average rates of half users are lower than 1.377bps/Hz for the spatially uncorrelated channel. As t increases, the rate reduces slowly. The reason is that spatial correlation results in cross paths between streams when there is more than one layer. If fading is strongly correlated in space, the inter-stream interference will be amplified which leads to a much lower data rate. Also, as t approaches 1, the rate witnesses a sudden drop. The reason is a large t results in a correlated R that increase the possibility of channel matrix $\mathbf{H}$ being rank-deficient. Since the number of available streams depends on the rank of $\mathbf{H}$, as t grows, there can be more correlated channel over which only one layer can be transmitted. An interesting fact is that comparing with the 1-

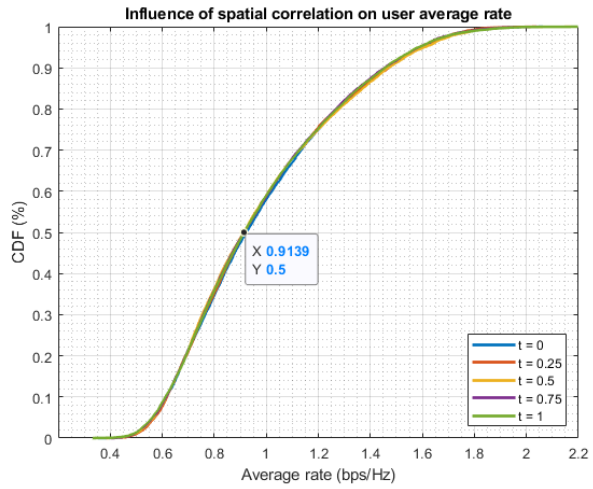

Fig. 10. Influence of the spatial correlation on user average rate ($n_r = 1$)

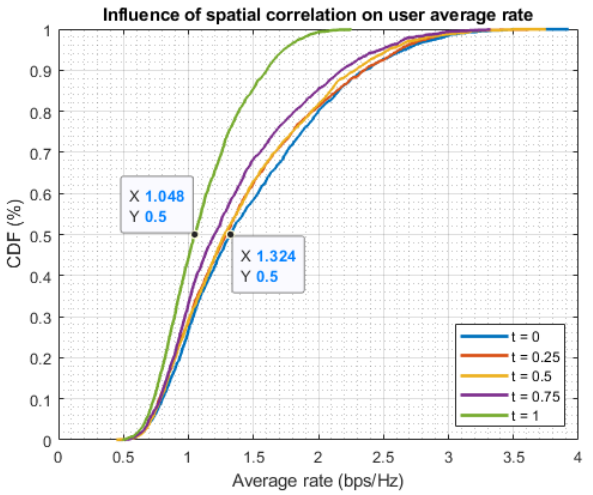

Fig. 11. Influence of the spatial correlation on user average rate ($n_r = 2$)

Rx case by Figure 10, the rate of the 2-Rx system is slightly larger even if $t = 1$. It means that more antennas always brings better performance, and even if the channel is strongly correlated, there can be some 2-stream case with CQI greater than single layer transmission. Moreover, for the 1-Rx system, or the 2-Rx case at low SINR that favours DET, only one stream is employed in transmission, and there is no inter-stream interference. Therefore, the gap between t = 0 and 1 is small at low SINR. As SINR increases, although influenced by inter-stream interference, the rate of 2-layer transmission can still be larger than that of 1-layer when the channel is good enough. To conclude, although the spatial correlation leads to inter-stream interference and reduces SINR, using more antennas can always provide a larger rate. Also, a larger spatial correlation results in more rank-deficient fading channels and less available streams, which restricts the diversity of transmission mode and data rate.

IV. Conclusion

We investigated the rate performance of LTE 4Tx SU-MIMO over a spatially and temporally correlated channel with quantised precoding and PF scheduling. It was proved that using more antennas brings more options in transmission mode and always benefits the rate, and a suitable scaling time scale can achieve a balance between fairness and performance. Also, the spatial and temporal correlation of fading results in a lower rate, although the influence is insignificant for most cases in practice. The major limitation is that the number of drops X is not large enough to produce smooth curves. With the code in Appendix, the simulation can be extended to more complicated cases.

V. Appendix: MATLAB code

The source code can be retrieved from https://github.com/SnowzTail/.

References

[1] R. T. Becker, "Precoding and spatially multiplexed mimo in 3gpp long-term evolution," *High Freq Electron*, 2009.
[2] B. Clerckx, "System-level performance evaluation of lte 4tx mimo downlink," January 2019.
[3] ——, "Wireless communications," January 2019.