Reinforcement Learning Aided Link Adaptation for Downlink NOMA Systems With Channel Imperfections

Qu Luo[*], Zeina Mheich[*], Gaojie Chen[*], Pei Xiao[*], and Zilong Liu[†]
[*] University of Surrey, Surrey, UK,
[†] University of Essex, Colchester, UK,
Email: q.u.luo@surrey.ac.uk, zeinamheich@hotmail.com, gaojie.chen@surrey.ac.uk,
p.xiao@surrey.ac.uk, and zilong.liu@essex.ac.uk.

Abstract—**Non-orthogonal multiple access (NOMA) is a promising candidate radio access technology for future wireless communication systems, which can achieve improved connectivity and spectral efficiency. Without sacrificing error rate performance, link adaptation combining with adaptive modulation and coding (AMC) and hybrid automatic repeat request (HARQ) can provide better spectral efficiency and reliable data transmission by allowing both power and rate to adapt to channel fading and enabling re-transmissions. However, current AMC or HARQ schemes may not be preferable for NOMA systems due to the imperfect channel estimation and error propagation during successive interference cancellation (SIC). To address this problem, a reinforcement learning based link adaptation scheme for downlink NOMA systems is introduced in this paper. Specifically, we first analyze the throughput and spectrum efficiency of NOMA system with AMC combined with HARQ. Then, taking into account the imperfections of channel estimation and error propagation in SIC, we propose SINR and SNR based corrections to correct the modulation and coding scheme selection. Finally, reinforcement learning (RL) is developed to optimize the SNR and SINR correction process. Comparing with a conventional fixed look-up table based scheme, the proposed solutions achieve superior performance in terms of spectral efficiency and packet error performance.**

Index Terms—**Non-orthogonal multiple access (NOMA), adaptive modulation and coding (AMC), hybrid automatic repeat request (HARQ), reinforcement learning (RL).**

I. Introduction

Recently, non-orthogonal multiple access (NOMA) has been envisioned as a promising multiple access technique to support diverse traffics with much stringent requirements, such as high spectral efficiency (SE) and high- level quality of service (QoS), for the beyond fifth generation (B5G) and sixth-generation (6G) network [1], [2]. Compared with the conventional orthogonal multiple access (OMA), NOMA can provide higher SE [2]. Existing NOMA techniques can be mainly categorized into two classes: power-domain NOMA [3] and code-domain NOMA (CD-NOMA) [4], [5]. This paper focuses on the power domain NOMA, which multiplexes users' signals in the power domain by superposition coding at the transmitter and employs successive interference cancellation (SIC) to decode the message at the receiver. Link adaptation combined with adaptive modulation and coding (AMC) and hybrid automatic repeat request (HARQ) is an another powerful technique to improve the system SE under an error performance constraint by dynamically adapting the code rate and the modulation order to the instantaneous fading channel condition [6]. To do that, the receivers periodically send channel quality indicator (CQI), via a feedback channel, to the transmitter to select an appropriate modulation and coding scheme (MCS). The mapping from instantaneous channel conditions to CQI is usually fixed to achieve a target reliability which is predicted through link-level simulations using a mathematical channel model.

To augment NOMA advantages, AMC has been widely considered [1], [3], [7]–[9]. A joint power allocation and AMC algorithm for downlink NOMA was developed in [1] to improve the SE and user fairness. The authors in [7] investigated the throughput performance of single-packet and HARQ with blanking for NOMA systems. Moreover, the authors in [9] proposed a resource allocation algorithm to tackle the problem of fair resource allocation and AMC for downlink NOMA systems. Different from [1], [7], [9], the block error rate requirement for each user was considered as a constraint in [8] when optimizing the AMC scheme. For uplink NOMA systems, an asymmetric adaptive modulation framework was introduced in [3] to address the distinct uncertainty of the bit error rate (BER) and throughput.

Besides the above valuable works, most existing AMC-NOMA schemes generally assume idea channel model, perfect channel estimation and SIC [1], [3], [7]–[9]. However, it is very difficult to obtain a high precision channel estimation due to many factors such as the imperfections of the transmitter and receiver circuits, which will also lead to error propagation during SIC [10]. In addition, the time-varying aspect of the channel and the delay in the feedback channels will also lead to the inaccuracy of the mathematical channel model [11]. The inaccuracy of the mathematical channel model will significantly degrades the system performance and affect users' QoS of AMC-NOMA. To tackle this issue, this paper proposes a reinforcement learning (RL) based algorithm to optimize the selection of MCS without relying on an explicit channel model. The main contributions of this paper are:

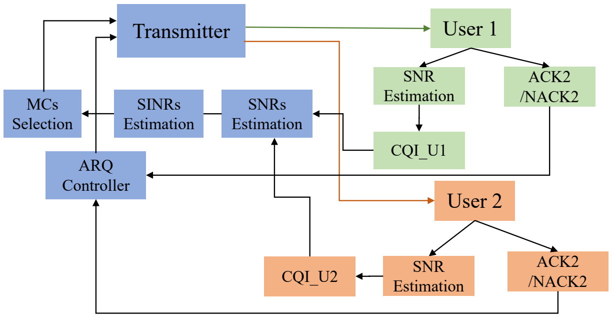


Fig. 1: System model of two-user NOMA with link adaptation.

- We analyze the average throughput and SE of the NOMA system with cross-layer scheme combining AMC and HARQ under the assumption of perfect SIC.
- We consider a more realistic setting where the MCS selection is inaccurate due to the imperfections in the system, we propose signal-to-noise ratio (SNR) and signal-to-interference-noise-ratio (SINR) based corrections to correct MCS selection.
- We propose a RL based algorithm to optimize the SNR and SINR corrections to ensure a maximum throughput fairness under a packet error rate (PER) performance constraint. up table based method.

II. Link adaptation aided NOMA system

A. Fundamental of NOMA

We consider a downlink NOMA system that consists of a base station (BS) that serves two users. Without loss of generality, the two users are denoted as U_1 and U_2. We further let x_1 and x_2 be the signals of U_1 and U_2 with a unit power, respectively. The transmitted signal at BS is the result of the superposition of the NOMA users' signals with appropriate power levels. Let α be the power fraction allocated for user U_1, then the received signal for $U_j, \forall j \in \{1,2\}$ is given by

$$y_j = h_j \left(\sqrt{\alpha} x_1 + \sqrt{1-\alpha} x_2\right) + n_j, \tag{1}$$

where $n_j, j = 1,2$ is the additive white Gaussian noise with zero mean and variance N_0, and h_j is the complexity channel coefficient between jth user and BS. The channel is assumed to be Rayleigh block fading, i.e., remains constant for the duration of one block or time slot, and changes independently between time slots. Denote $g_j = |h_j|^2/N_0$ as the channel SNR of jth user. Without loss of generality, we assume that the channels are ordered such that $g_1 \geq g_2$. At user side, SIC is performed to detect the transmitted message. Specifically, the weak user (U_2) directly decodes its signal by treating U_1 as noise, while the strong user (U_1) first decodes the signal of U_2 and then subtracted it from the received signal before decoding its own signal. As a result, the SINRs of the two users are respectively given as

$$\gamma_1 = \alpha g_1, \quad \gamma_2 = \frac{(1-\alpha) g_2}{1+\alpha g_2}, \tag{2}$$

under the assumption that there is no error propagation during SIC.

TABLE I: An example of MSC table.

TMs	TM0	TM1	TM2	TM3	TM4	TM5
Modulation	-	BPSK	QPSK	QPSK	16-QAM	16-QAM
Code rate	0	1/2	2/3	5/6	2/3	5/6

B. System model of NOMA with link adaptation

To enhance SE, a cross-layer scheme combining AMC with HARQ is incorporated for NOMA system, which is shown in Fig. 1. The main processes are summarized as follows:

- **Step 1:** The BS estimates the SNR from the reported CQIs of U_1 ans U_2 via the radio resource control signalling, denoted as CQI_U$_1$ and CQI_U$_2$, respectively. The CQI is a binary representation of the communication channel quality calculated at the receiver side.
- **Step 2:** The BS estimates the SINR $\gamma_j, j = 1,2$ according to (2).
- **Step 3:** The estimated SINRs are then mapped to corresponding MCSs for U_1 and U_2, denoted by TM_U1 and TM_U2, respectively. We assume U_1 and U_2 use a same MCS table.
- **Step 4:** The BS re-transmits the previous packet to user if a non-acknowledgement (NACK) is received when the number of re-transmissions is less than the maximum re-transmission. Otherwise, new data packet will be sent with the selected MCS.
- **Step 5:** After receiving the data packet, the user will attempt to decode the packet. If the packet is successfully decoded, the user will feed back an acknowledgement (ACK) to the BS. Otherwise, a NACK will be sent by the user.
- **Step 6:** User estimates channel condition and calculates SNR based on channel estimation value.
- **Step 7:** User maps the estimated SNR to corresponding CQI.
- **Step 8:** User reports CQI and ARQ information to BS.

The AMC-NOMA system has N transmission modes (TMs), each of them consists of a MCS. Table I shows an example of MCS table with $N = 6$ TMs. We assume the entire SINR range are divided into N non-overlapping consecutive intervals with switching thresholds denoted by $\{\theta_n\}_{n=0}^{N}$, i.e.,

$$\text{TM } n \text{ is chosen, when } \gamma_j \in [\theta_n, \theta_{n+1}). \tag{3}$$

In general, we have $\theta_N = +\infty$, and $\{\theta_n\}_{n=1}^{N}$ are determined to achieve a target PER constraint, denoted as P_{tar}, and better SE.

Remark 1: Obviously, the imperfections of the channels will lead to error propagation during SIC and affect the accuracy of SNR calculation, which further makes CQI unreliable at receiver and BS. Finally, the BS makes unwise scheduling and MCS decision based on the imperfect SINR, which reduces the system SE and users' QoS.

C. The average SE of AMC-NOMA with HARQ

The expected throughput performance of the TM n for the AMC-NOMA with HARQ is given by [12], [13]

$$\eta_n(\gamma) = R_n \cdot \log 2(M_n) \cdot \frac{1 - \prod_{i=0}^{N_r} \mathrm{PER}_{n,i}(\gamma)}{1 + \sum_{i=0}^{N_r-1} \prod_{j=0}^{i} \mathrm{PER}_{n,j}(\gamma)}, \tag{4}$$

where R_n and M_n are the code rate and modulation order of TM n, respectively, N_r is the number of HARQ re-transmissions, and $\mathrm{PER}_{n,i}(\gamma)$ is the PER of TM n on the ith re-transmission as a function of γ. The analytical expression of $\mathrm{PER}_{n,i}(\gamma)$ for can be approximated using curve fitting method [14], i.e.,

$$\mathrm{PER}_{n,i}(\gamma) \approx \begin{cases} 1, & 0 \leq \gamma < \gamma_{p_{n,i}} \\ b_{n,i} e^{-c_{n,i}\gamma}, & \gamma \geq \gamma_{p_{n,i}} \end{cases} \tag{5}$$

where the parameters $b_{n,i}$ and $c_{n,i}$ are the constants depend on the system settings, such as channel codes, constellation, etc. As a result, the average SE can be written as

$$\eta = \sum_{n=0}^{N-1} \int_{\gamma_n}^{\gamma_{n+1}} \eta_n(\gamma) f(\gamma) d\gamma, \tag{6}$$

s where $f(\gamma)$ is the probability density function (PDF) of the SINR γ. Assume that the channel SNRs g_1, g_2 are sampled from an exponential distribution $p_g = \frac{1}{\hat{g}} \exp\left(-\frac{g}{\hat{g}}\right)$ such that $g_1 \leq g_2$, where $\hat{g}$ is the average SNR for all user channels. Its cumulative density function (CDF) is $F_G(g) = 1 - \exp\left(-\frac{G}{\hat{g}}\right)$. According to the order statistics theory, the PDF of the ordered g_j is given by [15]

$$f_{g_j}(x) = \frac{J!}{(j-1)!(J-j)!} \left[F_G(x)\right]^{j-1} \left[1 - F_G(x)\right]^{J-j} p_x. \tag{7}$$

In (6), $f(\gamma)$ is the PDF of the SINR in (2), which depends on the channel SNRs and the power allocation factors α. By expressing the channel SNRs as function of the SINR for each user, $g_1 = \frac{\gamma_1}{\alpha}$ and $g_2 = \frac{\gamma_2}{1-\alpha-\alpha_2\gamma_2}$, the PDFs of the SINRs for the two users' NOMA systems are given as

$$\begin{aligned} f_{\gamma_1}(x) =& \frac{1}{\alpha} f_{G_1}\left(\frac{\gamma_1}{\alpha}\right), \\ f_{\gamma_2}(x) =& \frac{1-\alpha}{\left(1-\alpha-\alpha_2\gamma_2\right)^2} f_{G_2}\left(\frac{\gamma_2}{1-\alpha-\alpha_2\gamma_2}\right). \end{aligned} \tag{8}$$

By substituting (8) and 4 into (6), we can obtain the average SE for AMC-NOMA system. However, the above analysis does not take into account some realistic assumptions, such as error propagation in SIC, channel estimation errors with unknown models, etc. If we ignore them in the system design, then the real PER can be greater than the target one resulting in a loss in SE. To address this problem, we propose in the next section a RL aided link adaptation for NOMA systems to ensure that the PER is always lower than the target value while maximizing the SE of the NOMA users.

III. The proposed Reinforcement learning aided link adaptation

In this section, we investigate the AMC-NOMA system when presents the channel estimation errors and error propagation during SIC. Specifically, we propose SINR and SNR based corrections to correct the selection of transmission model and CQI, respectively. Then, a RL-based learning algorithm is proposed to optimize the correction process by maximizing the system SE subject to a target PER constraint.

A. SNR and SINR estimation

We assume that the transmitter estimates the channel gain for each user using the CQI received from it. The SNR estimated at the transmitter is equal to the SNR threshold corresponding to this CQI, corrupted by a random noise. In other terms, it is the minimum SNR required to guarantee the target PER for the TM number indicated by CQI. We assume that the transmitter has an MCS table used to map the SINR into MCS but this mapping is not necessarily perfect due to the inaccuracy of the assumed channel model used to determine the SNR thresholds. Thus the estimated SNR for user u_j, when its CQI is equal to i is

$$\hat{g_j} = \theta_i \epsilon_j, \tag{9}$$

where ϵ_j is a channel estimation error whose statistical model is unknown for both the transmitter and the receiver. It includes the errors/delays coming from multiple sources such as the errors in the CQI feedback channel, the delay of the feedback channels, the deviation from the assumed channel model, the imperfections of the SNR measurement unit, the systematic errors which cause bias, etc.

After estimating the SNR for each NOMA user, the transmitter calculates the SINR for each user as shown in (2) to determine its TM. In this case, the SINR for the two users are given by

$$\gamma_1 = \alpha\hat{g}_1, \quad \text{and} \quad \gamma_2 = \frac{(1-\alpha)\hat{g}_2}{1+\alpha\hat{g}_2}, \tag{10}$$

resepctively. For the sake of achieving user fairness, the transmitter chooses α such that the SINRs for both users are equal:

$$\alpha = \frac{\sqrt{(\hat{g}_1 + \hat{g}_2)^2 + \hat{g}_1\hat{g}_2^2}}{2\hat{g}_1\hat{g}_2}. \tag{11}$$

Based on the values of γ_1 and γ_2, the transmitter chooses the TMs for both users using the MCS table. The TMs for user 1 and user 2 are denoted by TM1 and TM2, respectively. The SNR estimation error as well as the error propagation in SIC both affect the choice of the TMs. They can lead to PER greater than the target one. In the following, we propose two methods to improve the spectral efficiency of the NOMA users in the presence of estimation errors. The first method consists of correcting the NOMA SINR while the second consists of correcting the channel SNR.

B. The proposed SINR correction

The SINRs in (10) estimated by the transmitter are corrupted by the SNR estimation error and do not take into account the error propagation in SIC. We propose to correct each SINR by multiplying it by a correction factor as follows:

$$\gamma_1^c = \alpha \hat{g}_1 \delta_1, \quad \gamma_2^c = \frac{(1-\alpha)\hat{g}_2}{1+\alpha\hat{g}_2}\delta_2, \tag{12}$$

where δ_1 and δ_2 are the correction factors for the first and second user SINR respectively. Based on the corrected SINR values γ_1^c and γ_2^c, the transmitter chooses the TMs for both users TM1 and TM2. The correction factors δ_1 and δ_2 should be optimized to maximize the minimum SE of user 1 and 2 under QoS constraints. We propose later to use RL algorithm in order to optimize the correction factors. Since in RL algorithm we need to discretize the space of the continuous variables δ_1 and δ_2, we consider an alternative approach to (12), based on correcting the TM instead of the SINR as follows:

$$\begin{aligned} \text{TM\_U1}^c &= q\left(\text{TM\_U1} + \delta_1\right), \\ \text{TM\_U2}^c &= q\left(\text{TM\_U2} + \delta_2\right), \end{aligned} \tag{13}$$

where q is a truncation function to ensure that the resulting TM belongs to the set of admissible values and TM_U1 and TM_U2 are the estimated TMs for U_1 and user U_2 respectively, using the erroneous SINRs in (10). In this case, δ_1 and δ_2 are discrete and take their values from the finite set $\{0, \pm 1, \ldots \pm (N-1)\}$.

C. The proposed SNR correction

Instead of correcting the SINR, the second method to improve the SE is to correct the estimated channel SNRs as follows:

$$g_1^c = \hat{g}_1 \delta_1, \quad g_2^c = \hat{g}_2 \delta_2. \tag{14}$$

Then, the BS calculates the SINR for each user using the corrected SNR values and determines the TMs for each user. Similar to SINR correction method, we consider an alternative approach for the SNR correction method based on correcting the CQI received from each user:

$$\begin{aligned} \text{CQI\_U1}^c &= q\left(\text{CQI\_U1} + \delta_1\right), \\ \text{CQI\_U2}^c &= q\left(\text{CQI\_U2} + \delta_2\right), \end{aligned} \tag{15}$$

in which the correction values are discrete. After correcting the CQIs, the transmitter uses the corrected CQI to estimate the channel SNRs according to (9) and calculates the SINRs which are used to determine the TM for the users.

D. RL-based link adaptation algorithm

RL is an area of machine learning which is about taking suitable action in an environment to maximize the reward in a particular situation, see e.g. [16]. It involves an agent, a set of states $\mathcal{S}$, and a set of actions per state $\mathcal{A}$. When the agent performs an action $a \in \mathcal{A}$, it makes a transition from state to state and receives a reward. The goal of the agent is to maximize its long-term reward by optimizing the action to be chosen in each state. Q-learning is a model-free RL algorithm (i.e., it does not require a model of the environment) to learn a function, called a policy that specifies the action to be taken by the agent which is in a certain state. Therefore, it has a Q-function that calculates the quality of each state-action combination:

$$Q : \mathcal{S} \times \mathcal{A} \rightarrow \mathbb{R}. \tag{16}$$

At each time t, after the agent selects an action $a_t \in \mathcal{A}$, it receives a reward r_t and moves from the state s_t to a state s_{t+1}. After that, Q is updated as follows

$$\begin{aligned} Q(s_t, a_t) \leftarrow & Q(s_t, a_t) \\ & + \beta \left(r_t + \rho \cdot \max_a Q(s_t, a) - Q(s_t, a_t) \right), \end{aligned} \tag{17}$$

where $\beta \in [0,1]$ is the learning rate and $\rho \in [0,1]$ is the discount factor. The initial values of Q are initialized to zero.

In the proposed NOMA system with link adaptation, the model of the estimation errors is not known at the transmitter and the receiver and is unpredictable. Hence, we propose to use RL algorithm to optimize the correction factors as function of the CQI received from the receivers. We define the state space, the action space and the reward as follows:

- The states s consist of the CQIs received from U_1 and U_2: $s = (\text{CQI\_U1}, \text{CQI\_U2})$.
- The actions a consist of the correction factors that should be optimized for each state value: $a = (\delta_1, \delta_2)$, $\delta_1, \delta_2 \in \{0, \pm 1, \ldots \pm (N-1)\}$.
- The receiver of each user sends an ACK or NACK message to the sender depending on whether the decoding of the packet as successful or not. The reward $\mathcal{R}$ calculated by the transmitter is equal to the minimum between the instantaneous throughput of the two users:

  $$\mathcal{R} = \min\{R_1 \log_2(M_1)\kappa_1, R_2 \log_2(M_2)\kappa_2\}, \tag{18}$$

  where $\kappa_j = 0$ if u_j sends a NACK and $\kappa_j = 1$ if u_j sends a ACK.

At each time t when the agent is in state s_t, the action a_t is selected to perform either exploration (randomly) or exploitation (the action a which maximize $Q(s_t, a)$). We choose to use the simplest approach, called ε-greedy, where $0 < \varepsilon < 1$ is a parameter controlling the amount of exploration and exploitation. The value of ε is decreased with time to make the agent exploits more. The reward defined above does not take into account the average PER. Although the MCS thresholds have been optimized to meet the target PER, i.e., P_{tar}, the real measured PER can be greater than the target one due to the errors and the imperfections in the SINR estimation. The transmitter estimates the current average PER at a certain episode of the Q-learning algorithm using the NACK messages as follows:

$$\hat{\text{PER}} = \frac{N_{\text{NACK}}}{N_p}, \tag{19}$$

where N_p is the total number of transmitted packets and N_{NACK} is the number of unsuccessfully decoded packets. To ensure that the PER is less than the target one P_{tar}, we set a negative reward R^- when $\hat{\text{PER}}$ is greater than P_{tar}, where

TABLE II: MCS table in NOMA system.

TM(n)	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
R	0	0.2	0.25	0.3	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8
M	-	2	2	2	4	4	4	4	4	4	4	4	4	4	4	4	4	16	16	16	16	16	16	16	16
SE	0	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.8	2.0	2.2	2.4	2.6	2.8	3.0	3.2
10% FER SNR (θ_n) [dB]	-	-5.4	-4.5	-3.6	-2.9	-1.9	-1.1	-0.4	0.40	1.00	1.60	2.16	2.74	3.34	3.91	4.46	5.24	6.35	7.21	8.21	9.20	10.15	11.17	12.35	14.45

Algorithm 1 RL-based link adaptation algorithm.

```
Initialize $Q(s,a) = 0, \forall a \in \mathcal{A}, s \in \mathcal{S}$.
for episode $t \leftarrow 1$ to $T$ do
    $U_1$ and $U_2$ perform data detection and decoding based on SIC and corresponding channel code, and estimate their SNRs
    Obtain $s_t = (\text{CQI\_U1}, \text{CQI\_U2})$ and ARQ information, i.e, $\kappa_{1,t}$ and $\kappa_{2,t}$
    Calculate $\mathcal{R}_t$ based on (18)
    if $t \geq N_{\text{th}}$ then
        Calculate $\hat{\text{PER}}_j = \frac{N_{j,\text{NACK}}}{N_p}$
    end if
    if $\hat{\text{PER}}_j > \text{PER}_{\text{tar}}$ then
        $\mathcal{R}_t = R^-$
    end if
    Determine the action $(\delta_1, \delta_2) = \arg\max_a Q(s_t, a)$ with a probability of $1 - \epsilon$, otherwise, select a random action
end for
return All $Q(s,a)$
```

R^- tuned by the transmitter. The pseudocode of the proposed RL based link adaptation for downlink NOMA is provided in Algorithm 1.

IV. Numerical results

We employ SIC for data detection and choose the low-density parity check codes (LDPC) with frame length equal to 200 bits from the 3GPP standard for 5G new radio [1] for channel encoding/decoding. We consider the MCS table in Table II, where the SNR thresholds $\theta_n, n = 1, 2, \ldots, 25$ are calculated to achieve a target FER less or equal to $P_{\text{tar}} = 0.1$ using the above LDPC codes with perfect channel estimations. As can be seen in Table II, there are $N = 25$ TMs, and the SNR threshold ranges from -5.39 dB to 14.45 dB. The maximum re-transmission number is $N_r = 3$ and N_{th} is set to be $N_{\text{th}} = 200$ for a reliable average PER estimation. The learning rate and discount factor are set to be $\beta = 0.9$ and $\rho = 0.1$, respectively. An ε-greedy policy with a fixed value of $\varepsilon = 0.08$ is employed during the learning phase. The goal is to have the RL agent performs in the long run.

The baseline solutions are the fixed look-up table, which is given in Table II, refereed to as fixed AMC. In the fixed look-up table approach, a static mapping of SINR to MCS is obtained by analyzing the PER curves and selecting the best MCS that satisfies the target PER. For example, if the received SNR is 3 dB, respectively, the transmitters will choose the TMs 12.

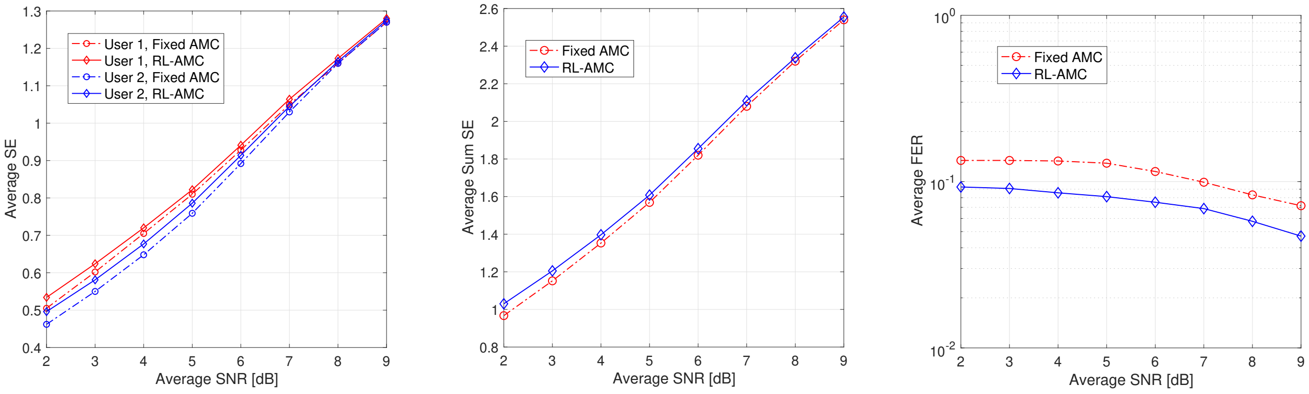
We first evaluate the average SE, average sum SE and average FER of SINR based correction, which is shown in Fig. 2. The channel estimation errors for the two users are modeled as $\epsilon_1 \sim \mathcal{N}(0.3, 0.3)$ and $\epsilon_2 \sim \mathcal{N}(0.3, 0.5)$, and $\mathcal{N}(m, v)$ denotes a Gaussian distribution with mean m and variance v. It is noted that the distribution of the channel estimation errors are not known for the transmitter. Without the aid of RL, we observe that the average FER can exceed the target one due to the errors in the channel estimation that makes the SNR thresholds θ_n inaccurate. As a result, the average SE decreases. With the aid of RL, we observe that the average FER is always less than the target one and the average SE is improved.

We further evaluate the performance when the average of the channel estimation error increases. Specifically, we plot in Fig. 3 the average SE, average sum SE and the average FER as function of $\hat{g}$, when the channel estimation errors for the two users are modeled as $\epsilon_1 \sim \mathcal{N}(1.2, 0.3)$ and $\epsilon_2 \sim \mathcal{N}(1.2, 0.3)$. We observe in Fig. 3(a) that the SE of the user having the best channel is greatly reduced due to two factors: the estimation error of its channel SNR and the error propagation in SIC due to the high probability of choosing a wrong TM for the second user. With the aid of RL, the transmitter is able to correct the TMs for both user resulting in improved SE and lower FER for both users. The performance of RL-aided link adaptation is shown for SINR correction and SNR correction methods. Both methods achieve similar average sum SE and are able to meet the target FER.

V. Conclusion

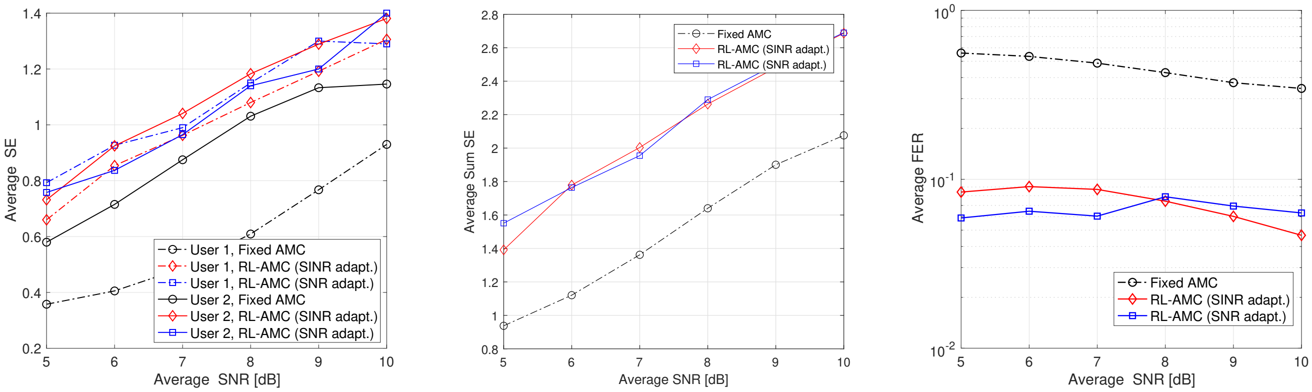
In this paper, we introduced a RL based link adaptation for NOMA systems. We analysed the throughput performance and average SE of a two-user NOMA system with AMC combining HARQ scheme conditioned on the perfect channel information and SIC. To address the inaccurate MCS selection due to the imperfections of SIC and channel estimation in the system, we proposed two approaches to correct the received

[1] http://www.3gpp.org/ftp//Specs/archive/38 series/38.212/



(a) Average SE (b) Average sum SE (c) Average FER

Fig. 2: Performance evaluation $\epsilon_1 \sim \mathcal{N}(0.3, 0.5)$ and $\epsilon_2 \sim \mathcal{N}(0.3, 0.8)$.



(a) Average SE (b) Average sum SE (c) Average FER

Fig. 3: Performance evaluation for $\epsilon_1 \sim \mathcal{N}(1.2, 0.3)$ and $\epsilon_2 \sim \mathcal{N}(1.2, 0.3)$.

MCS selection, i.e., SNR and SINR based corrections. Finally, RL algorithm was developed to optimize the SNR and SINR correction process. Numerical results demonstrated the benefits of the proposed RL-based link adaptation scheme in terms of SE and FER.

References

[1] W. Yu, H. Jia, and L. Musavian, "Joint adaptive M-QAM modulation and power adaptation for a downlink NOMA network," *IEEE Trans. Commun.*, vol. 70, no. 2, pp. 783–796, 2021.

[2] Q. Luo *et al.*, "An error rate comparison of power domain non-orthogonal multiple access and sparse code multiple access," *IEEE Open J. Commun. Soc.*, vol. 2, no. 4, pp. 500–511, Mar. 2021.

[3] K. Wang, T. Zhou, T. Xu, H. Hu, and X. Tao, "Asymmetric adaptive modulation for uplink NOMA systems," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7222–7235, 2021.

[4] Q. Luo, Z. Liu, G. Chen, Y. Ma, and P. Xiao, "A novel multitask learning empowered codebook design for downlink SCMA networks," *IEEE Wireless Commun. Lett.*, vol. 11, no. 6, pp. 1268–1272, 2022.

[5] Q. Luo *et al.*, "A novel non-coherent SCMA with massive MIMO," *IEEE Wireless Commun. Lett.*, vol. 11, no. 11, pp. 2250–2254, 2022.

[6] 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, 2004.

[7] Z. Mheich, W. Yu, P. Xiao, A. U. Quddus, and A. Maaref, "On the performance of harq protocols with blanking in NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 19, no. 11, pp. 7423–7438, 2020.

[8] H. Yahya, E. Alsusa, and A. Al-Dweik, "Design and analysis of NOMA with adaptive modulation and power under BLER constraints," *IEEE Trans. Veh. Techno.*, pp. 1–6, 2022.

[9] H.-Y. Hsieh, M.-J. Yang, and C.-H. Wang, "Fair resource allocation using the mcs map for multi-user superposition transmission (MUST)," in *2016 IEEE 27th PIMRC*. IEEE, 2016, pp. 1–7.

[10] N. P. Le and K. N. Le, "Uplink NOMA short-packet communications with residual hardware impairments and channel estimation errors," *IEEE Trans. Vehi. Techno.*, vol. 71, no. 4, pp. 4057–4072, 2022.

[11] S. Schiessl, M. Skoglund, and J. Gross, "NOMA in the uplink: Delay analysis with imperfect CSI and finite-length coding," *IEEE Trans. Wireless Commun.*, vol. 19, no. 6, pp. 3879–3893, 2020.

[12] R. Sassioui, M. Jabi, L. Szczecinski, L. B. Le, M. Benjillali, and B. Pelletier, "HARQ and AMC: Friends or foes?" *IEEE Trans. Commun.*, vol. 65, no. 2, pp. 635–650, 2016.

[13] P. Zhang, Y. Miao, and Y. Zhao, "Cross-layer design of AMC and truncated HARQ using dynamic switching thresholds," in *2013 WCNC*. IEEE, 2013, pp. 906–911.

[14] J. Ramis and G. Femenias, "Cross-layer design of adaptive multirate wireless networks using truncated harq," *IEEE Trans. Veh. Techno.*, vol. 60, no. 3, pp. 944–954, 2011.

[15] H. A. David and H. N. Nagaraja, *Order statistics*. John Wiley & Sons, 2004.

[16] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.