Robust Adaptive Modulation and Coding (AMC) Selection in LTE Systems using Reinforcement Learning

Raffaele Bruno, Antonino Masaracchia, Andrea Passarella
Institute of Informatics and Telematics (IIT)
Italian National Research Council (CNR)
Via G. Moruzzi 1, Pisa, ITALY
E-mail: {r.bruno, a.masaracchia,a.passarella}@iit.cnr.it

***Abstract*—Adaptive Modulation and Coding (AMC) in LTE networks is commonly employed to improve system throughput by ensuring more reliable transmissions. Most of existing AMC methods select the modulation and coding scheme (MCS) using pre-computed mappings between MCS indexes and channel quality indicator (CQI) feedbacks that are periodically sent by the receivers. However, the effectiveness of this approach heavily depends on the assumed channel model. In addition CQI feedback delays may cause throughput losses. In this paper we design a new AMC scheme that exploits a reinforcement learning algorithm to adjust at run-time the MCS selection rules based on the knowledge of the effect of previous AMC decisions. The salient features of our proposed solution are: i) the low-dimensional space that the learner has to explore, and ii) the use of direct link throughput measurements to guide the decision process. Simulation results obtained using ns3 demonstrate the robustness of our AMC scheme that is capable of discovering the best MCS even if the CQI feedback provides a poor prediction of the channel performance.**

***Index Terms*—LTE, channel quality, adaptive modulation and coding (AMC), reinforcement learning, performance evaluation.**

I. Introduction

The Long Term Evolution (LTE) is an acronym that refers to a series of cellular standards developed by 3GPP to meet the requirements of 4G systems. In particular, LTE has been designed to provide high data rates, low latency, and an improved spectral efficiency compared to previous cellular systems. To achieve these goals LTE adopts advanced physical layer technologies, such as OFDMA and multi-antenna techniques, and it supports new Radio Resource Management (RRM) functions for link adaptation [1]. In particular, adaptive modulation and coding (AMC) has been proposed for LTE, as well as many other wireless communication systems, to increase channel throughput [2]. In general, AMC techniques try to optimally select the channel coding and modulation scheme (MCS), while fulfilling a certain Block Error Rate (BLER) constraint[1] by taking into account the current channel conditions and the receiver's characteristics (e.g., antenna configuration). For LTE downlink transmissions, traditional AMC schemes rely on the channel quality indicator (CQI) feedbacks that are periodically reported by the user terminals (UEs) to their base stations (eNBs) [3]. How CQI values should be computed by the UE using channel state information (e.g., SINR measurements) is implementation dependent. In principle, an eNB can use other information in addition to the CQI values reported by UEs, such as HARQ retransmissions, to determine the selected MCS. In practical implementations - as better explained in Section II - the UEs directly selects the MCS value that, if used by the eNB under the measured channel conditions, would achieve the maximum possible throughput by guaranteeing that the BLER is below 10%. This value is then mapped onto a CQI value and fed back to the eNB (that translates it back into the corresponding MCS value) [4], [5]. Therefore, the key focus of AMC algorithms is to define how UEs can compute MCS values that satisfy the BLER requirements.

Several technical challenges have to be addressed to design efficient AMC solutions for LTE systems. In particular, in practical LTE systems, the SINR values of multiple subcarriers are aggregated and translated into a one-dimensional link quality metric (LQM), since the same MCS must be assigned to all subcarriers assigned to each UE. Popular methods that are used in LTE to obtain a single effective SINR from a vector of physical-layer measurements related to subcarriers are the exponential effective SINR mapping (EESM) [6] or the mean mutual information per coded bit (MMIB) [7]. Once the LQM is found, AMC schemes typically exploit *static mappings* between these link quality metrics and the BLER performance of each MCS to select the best MCS (in terms of link throughput). In other words, for each MCS a range of LQM values is associated via a look-up table, over which that MCS maximises link throughput. Either link-level simulations or mathematical models can be used to generate such static BLER curves under a specific channel model. Unfortunately, past research has shown that it is difficult to derive accurate link performance predictors under realistic channel assumptions [5], [8]–[10]. Furthermore, a simulation-based approach to derive the mapping between LQM values

[1]The BLER for a certain user is defined as the ratio between the number of erroneous resource blocks and the total number of resource blocks received by that user. In the LTE standard it is mandated that the selected MCS ensures an average BLER under the measured channel conditions lower than 10% [3].

and BLER performance is not scalable since it is not feasible to exhaustively analyse all possible channel types or several possible sets of parameters [11]. The second main problem with table-based AMC solutions is that a delay of several transmission time intervals (TTIs) may exist between the time when a CQI report is generated and the time when that CQI feedback is used for channel adaptation. This is due to processing times but also to the need of increasing reporting frequency to reduce signalling overheads. This mismatch between the current channel state and its CQI representation, known as *CQI ageing*, can negatively affect the efficiency of AMC decisions [12], [13]

To deal with the above issues, in this paper we propose a new flexible AMC framework, called RL-AMC, that autonomously and at run-time decides upon the best MCS (in terms of maximum link-layer throughput) based on the knowledge of the outcomes of previous AMC decisions. To this end we exploit reinforcement learning techniques to allow each eNB to update its MCS selection rules taking into account past observations of achieved link-layer throughputs. Specifically, the purpose of the decision-making agent in our AMC scheme is to discover which is the correction factor that should be applied to CQI feedbacks in order to guide the transmitters in selecting more efficient MCSs. An important feature of our proposed scheme is the use of a low-dimensional state space, which ensures a robust and efficient learning even under time-varying channel conditions and mobility. Through simulations in ns3 we show that our AMC method can improve the LTE system throughput compared to other schemes that use static mappings between SINR and MCS both under pedestrian and vehicular network scenarios. Furthermore, our AMC is capable of discovering the best MCS even if the CQI feedback provides a poor prediction of the channel performance.

Before presenting our solution, it is important to point out that other studies [14]–[17] have proposed to use machine learning techniques to improve AMC in wireless systems. The main weakness of most of these solutions is to rely on machine learning algorithms (e.g., pattern classification [15] or SVM [14], [16]) that require large sets of training samples to build a model of the wireless channel dynamics. Similar to our work, the AMC scheme proposed in [17] exploits Q-learning algorithms to avoid the use of model-training phases. However, the MCS selection problem in [17] is defined over a continuous state space (i.e., received SINR), and even after discretisation a large number of states must be handled by the learning algorithm.

The remaining of this paper is organised as follows. Section II overviews existing proposals to implement AMC techniques in LTE networks. Section III introduces the principles of reinforcement learning, and introduces the Q-learning algorithm. Section IV describes our RL-AMC scheme. In Section V we report simulation results to demonstrate the performance improvements of the proposed scheme. Section VI concludes the paper with final remarks.

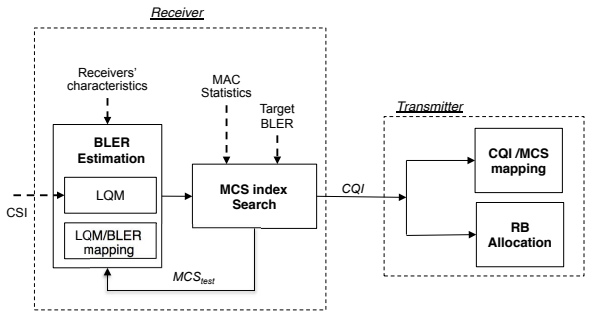

Fig. 1. AMC functional architecture.

II. AMC IN LTE

For the sake of illustrative purposes, in Figure 1 we show a functional architecture for a practical AMC scheme for LTE systems. At the receiver's side, a first module is responsible for processing the channel state information (e.g., per-subcarrier received SINR values) to obtain a BLER estimation under the assumption of a specific channel model. Specifically, the receiver maps the channel measurements into a single link quality metric. Then, an offline look-up table is used to map this LQM to a BLER estimate for each MCS. These BLER curves are used to find the highest-rate MCS index that can satisfy a 10% BLER target. Finally, the selected MCS index is sent in the form of a CQI feedback to the transmitter. Based on such CQI feedback the transmitter performs resource scheduling and MCS selection.

Most of existing research on AMC schemes for LTE is focused on the problem of CQI calculation given a link quality metric. As mentioned in Section I a popular and sufficiently accurate method for LQM calculation is EESM. For instance, the authors in [18] study the MCS performance under an AWGN channel. Accurate packet error prediction for link adaptation via a Gaussian approximation of coding and decoding performance is proposed in [19]. A novel LQM metric for link adaptation based on raw bit-error-rate, effective SINR and mutual information is investigated in [20]. In [4] the authors proposed MCS selection based on packet-level effective SINR estimates rather than block-level SINR values, and they describe different averaging schemes to map BLER onto packet error rates. On the other hand, the authors in [5], [21] develops statistical models of the EESM under different channel models and use those models to analyse the throughput of EESM-based AMC for various CQI feedback schemes. A second group of paper studies channel predictors to deal with the CQI ageing. The authors in [12] derive closed-form expressions for the average throughput of an adaptive OFDMA system under the assumption of imperfect CQI knowledge. The performance of different CQI predictors, such as Kalman filtering or linear prediction with stochastic approximation, are evaluated in [13] and [22].

III. Background on Reinforcement Learning (RL)

Reinforcement Learning (RL) is a popular machine learning technique, which allows an agent to automatically determine the optimal behaviour to achieve a specific goal based on the positive or negative feedbacks it receives from the environment in which it operates after taking an action from a known set of admissible actions [23]. Typically, reinforcement learning problems are instances of the more general class of Markov Decision Processes (MDPs), which are formally defined through:

- a finite set $S = \{s_1, s_2, \ldots, s_n\}$ of the n possible states in which the environment can be;
- a finite set $A(t) = \{a_1(t), a_2(t), \ldots, a_m(t)\}$ of the m admissible actions that the agent may perform at time t;
- a transition matrix P over the space S. The element $P(s, a, s')$ of the matrix provides the probability of making a transition to state $s' \in S$ when taking action $a \in A$ in state $s \in S$;
- a reward function R that maps a state-action pair to a scalar value r, which represents the immediate payoff of taking action $a \in A$ in state $s \in S$.

The goal of a MDP is to find a *policy* π for the decision agent, i.e., a function that specifies the action that the agent should choose when in state $s \in S$ to maximise its expected long-term reward. More formally, if an agent follows a policy π starting from a certain state s at time t the policy value over an infinite time horizon, also called the value-state function, is simply given by

$$V^{\pi}(s) = \sum_{k=0}^{\infty} \gamma^k r_{t+k} \ , \tag{1}$$

where $\gamma \in [0, 1]$ is a *discount factor* that weights future rewards. Then an *optimal* policy π^* is, by definition, the one that maximise the value-state function. As a consequence, the policy that ensures the maximum possible expected reward, say $V^*(s)$, could be obtained by solving an optimisation problem $V^*(s) = \max_{\pi} V^{\pi}(s)$. If the transition matrix is known such optimisation problem can be expressed using a system of nonlinear equations by using techniques such as dynamic programming [23]. However, in most practical conditions it is hard, if not even impossible, to acquire such complete knowledge of the environment behaviour. In this case there are model-free learning methods that continuously update the probabilities to perform an action in a certain state by exploiting the observed rewards. Such methods adopt an alternative characterisation of policy goodness based on the state-action value function, or Q-function. Formally, the function $Q^{\pi}(s, a)$ computes the expected reward of taking an action a in a starting state s and then following the policy π hereafter. Owing to the Bellman's optimality principle, it holds that a greedy policy (i.e., a policy that at each state selects the action with the largest Q-value) is the optimal policy. In other words, it holds that $V^*(s) = \max_{a \in A} Q^*(s, a)$ with $Q^*(s, a) = \max_{\pi} Q(s, a)$.

In this work we use a model-free solving technique for reinforcement learning problems known as *Q-learning* [24], which constructs the optimal policy by iteratively selecting the action with the highest value in each state. The core of this algorithm is an iterative value update rule that each time the agent selects an action and observes a reward makes a correction of the old Q-value for that state based on the new information. This updating rule is given by:

$$Q(s, a) = Q(s, a) + \alpha \left[r(s, a) + \gamma \max_{a'} Q(s', a') - Q(s, a) \right] \ , \tag{2}$$

where $\alpha \in [0, 1]$ is the learning rate. Basically, the α parameter determines the weight of the newly acquired information over state-action value information. In our AMC framework we use $\alpha = 0.5$.

The advantage of Q-learning is that it is guaranteed to converge to the optimal policy. On the negative side, the convergence speed may be slow if the state space is large due to the *exploration vs. exploitation dilemma* [23]. Basically, when in state s the learning agent should exploit its accumulated knowledge of the best policy to obtain high rewards, but it must also explore actions that it has not selected before to find out a better strategy. To deal with this issue, various exploration strategies have been proposed in the literature, ranging from simple greedy methods to more sophisticated stochastic techniques, which assign a probabilistic value for each action a in state s according to the current estimation of $Q(s, a)$. In Section IV we discuss more in detail such exploration strategies.

IV. An RL-based AMC Scheme (RL-AMC)

In order to apply the Q-learning approach to the MCS selection problem it is necessary to define: i) the state space of the problem, ii) the feedbacks that the decision agent receives from the LTE network, and iii) the admissible actions for the agent with the action selection strategy. In our RL-based AMC framework, the problem state consists of CQI feedbacks and their evolution trends. The reward is the instantaneous link throughput obtained by a user after each transmission. Finally, an action is the selection of a correction factor to be applied to each CQI feedback to identify the best MCS under the current channel conditions. In the following, we describe in details the operations of our proposed AMC algorithm.

First of all, it is important to clarify that the AMC decision agent interacts with the environment (i.e., the LTE network) at discrete time instants, called epochs. At each epoch the agent receives some representation of the LTE channel state and on that basis selects an action. In the subsequent epoch the agent receives a reward, and finds itself in a new state. In our AMC framework we assume that an epoch is the time when the UE receives a segment of data, either new or retransmitted. Without loss of generality we also assume that the decision agent is provided with a mapping rule that establishes a relationship between SINR values and MCS indexes. Note that our solution is not restricted to any specific BLER models but *an initial MCS value is only needed to bootstrap the*

learning process and to reduce the size of the state space. Thus, it is not necessary that this mapping is accurate nor adjusted to the unique characteristics of each communication channel. In Section V we will investigate the robustness of our AMC scheme to inaccurate CQI representation of channel performance.

Intuitively, a straightforward approach to define the state of the MCS selection problem would be to use the SINR values of received segments of data[2] as state variables, as in [17]. However, the SINR is a continuos variable and it should be discretised to be compatible with a discrete MDP formulation. The main drawback is that a fine discretisation leads to a large-dimensional state space, which increases convergence and exploration times. To avoid this problem, we directly use CQI-based metrics for the state representation. Specifically, we adopt a two-dimensional space $S=\{s_1, s_2\}$ to characterise the LTE communication channel. The first state variable represents the CQI value (called CQI^m) that the UE should select using the internal look-up table that associates BLER and MCS and received SINR. The second state variable represents the ΔCQI^m value, which is defined as the difference between the last two consecutive CQI^m estimates. In other words, ΔCQI^m provides a rough indication of the trend in channel quality evolution. For instance, $\Delta\text{CQI}^m < 0$ implies that the channel quality is temporarily degrading.

Since the objective of the MCS selection procedure should be to maximise the link throughput it is a natural choice to define the reward function as the instantaneous link-layer throughput achieved when taking action a (i.e., applying a correction factor to current CQI value taken from the mapping function) when in state s (i.e., given the pair $\{\text{CQI}_t^m, \Delta\text{CQI}_t^m\}$). More precisely, we assume that the reward value of an erroneous downlink transmission is null. On the other hand, the reward for a successful downlink transmission is given by

$$R(s_{t_1}, a_{t_1}) = \frac{TB}{\#TTIs \text{ in } [t_1, t_2]} \ , \tag{3}$$

where TB is the MAC transport block size (i.e., the number of useful bits that could be carried in a certain number of RBs with a certain MCS), while the denominator is the time between the time t_1 when that segment of data was first scheduled and the time t_2 when it was successfully received[3].

The core of the Q-learning algorithm is represented by the set A of admissible actions. In our learning model we assume that an action consists of applying a correction factor to the CQI value that is initially estimated by means of the internal look-up table. As discussed above, the mapping relationship between SINR values and MCS may be inaccurate and the correction factor allows the agent to identify the best modulation and coding scheme (in the sense of maximising the link throughput) for the given channel conditions. For instance, it may happen that the SINR-to-MCS mapping is too conservative for the current channel conditions and an MCS with an higher data rate can be used without violating the target BLER requirement. In this case the correction factor should be positive. Furthermore, a correction factor is also needed to compensate eventual errors due to CQI feedback delay. More formally, we assume that an action taken by the AMC decision agent at time t is one possible choice of an integer number in the set $(-k, \ldots, -2, -1, 0, 1, 2, \ldots k)$, that we denote as a_t in the following. This index is added to the original CQI^m value to compute the CQI to be sent to the eNB, denoted as CQI^f. The line of reasoning for this adjustment is as follows. Let us assume that the agent state at time t is $\{\text{CQI}_t^m, \Delta\text{CQI}_t\}$. We argue that if $\Delta\text{CQI}_t < 0$ we should prefer conservative MCS selections (and thus use values of a_t lower than 0) because the channel trend is negative, while if $\Delta\text{CQI}_t \geq 0$ we can try to use MCSs offering higher data rates (and thus positive values for a_t). Recalling that the CQI is an integer between 0 and 15 [3], this can be expressed by writing that the CQI feedback, say CQI_t^f, that should be sent to the eNB by the UE to guide the selection of the MCS index for downlink transmissions at next epoch $t+1$ should be

$$\text{CQI}_t^f = \max\left[0, \min\left[\text{CQI}_t^m + a_t, 15\right]\right] \ , \tag{4}$$

where $a \in [0, 1, 2, \ldots k]$ if $\Delta\text{CQI}_t \geq 0$ and $a \in [-k, \ldots, -2, -1, 0]$ otherwise. Thus, the set of admissible actions is different whether the channel-quality trend is negative or non-negative. Before proceeding it is useful to point out that the choice of the k value determines how aggressively we want to explore the problem state space. In general, the selection of the k value could take into account the CQI difference statistics, i.e., to what extent a current CQI may be different from the reported CQI after a feedback delay [10]. In Section V-C we will discuss this aspect more in detail.

A very important learning procedure is the action selection rule, i.e., the policy used to decide which specific action to select in the set of admissible actions. As discussed in Section III there is a tradeoff between exploitation (i.e., to select the action with the highest Q-value for the current channel state) and exploration (i.e., to select an action randomly). The simplest approach (called ϵ-greedy [23]) would be to use a fixed probability ϵ to decide whether to exploit or explore. A more flexible policy (called *softmax* action-selection rule [23]) is to assign a probability to each action, basing on the current Q-value for that action. The most common softmax function used in reinforcement learning to convert Q-values into action probabilities $\pi(s, a)$ is the following [23]:

$$\pi(s, a) = \frac{e^{Q(s,a)/\tau}}{\sum_{a' \in \Omega_t} e^{Q(s,a')/\tau}} \ , \tag{5}$$

where Ω_t is the set of admissible actions at time t. Note that for high τ values the actions tend to be all (nearly) equiprobable. On the other hand, if $\tau \to 0$ the softmax policy becomes the

[2]We recall that LTE physical layer relies on the concept of resource blocks. A segment of data or transport block is basically a group of resource blocks with a common MCS that are allocated to a user. Typically, a packet coming from the upper layers of the protocol stack will be transmitted using multiple segments of data.

[3]A segment of data that is discarded after a maximum number of retransmissions has also a null reward.

same as a merely greedy action selection. In our experiments we have chosen $\tau = 0.5$.

V. Performance Evaluation

In this section, we assess the performance of our proposed RL-AMC scheme in two different scenarios. In the first one a fixed CQI is fed back to the eNB by each UE. Without the use of reinforcement learning AMC necessarily selects a fixed MCS independently of the current channel conditions. Then, we demonstrate that our RL-based AMC is able to converge towards the best MCS even if the initial CQI estimate are totally wrong. In the second scenario we compare RL-AMC against the solution described in [25], which exploits spectral efficiency estimates to select MCS. Specifically, the spectral efficiency of user i is approximated by $\log_2(1+\gamma_i/\Gamma)$, where γ_i is the effective SINR of user i and Γ is a scaling factor. Then, the mapping defined in the LTE standard [26] is used to convert spectral efficiency into MCS indexes and, then, into CQI feedbacks. In this case, we show that our reinforcement learning algorithm is able to improve the accuracy of the CQI mapping at run time.

A. Simulation setup

All the following experiments have been carried out using the ns3 packet-level simulator, which includes a detailed implementation of the LTE radio protocol stack. As propagation environment, we assume an *Urban Macro* scenario, where path loss and shadowing are modelled according to the COST231-Hata model [27], which is widely accepted in the 3GPP community. The fast fading model is implemented using the Jakes model for Rayleigh fading [28]. To limit the computation complexity of the simulator pre-calculated fading traces are included in the LTE model that are based on the standard multipath delay profiles defined in [29]. In the following tests we have used the *Extended Typical Urban* fading propagation model with pedestrian (3 km/h) and vehicular (30 km/h) users' speeds. The main LTE physical parameters are summarised in Table I. Regarding the network topology, the considered scenario is composed by a single cell and a number of users, chosen in the range $[10, 100]$, which move according a Random Waypoint Model (RWM) [30] within the cell, if not otherwise stated. A downlink flow, modelled with an infinite buffer source, is assumed to be active for each UE. Finally, the eNode B adopts the resource allocation type 0, thus only allocating resource block groups (RBGs) to scheduled UEs. Given the downlink system bandwidth (see Table I) a RBG comprises two RBs [3]. RBGs are assigned to UEs following a Round Robin (RR) scheduler that divides equally the available RBGs to active flows. Then, all the RBs in the allocated RBGs used the MCS index that is signalled in the last received CQI feedback. Furthermore, the implemented version of RR algorithm is not adaptive, which implies that it maintains the same RBGs and MCS index when allocating retransmission attempts.

All results presented in the following graphs are averaged over five simulation runs with different network topologies.

TABLE I
Simulation parameters.

Parameter	Value
Carrier frequency	2GHz
Bandwidth for downlink	5 MHz
eNB power transmission	43 dBm
Subcarrier for RB	12
SubFrame length	1 ms
Subcarrier spacing	15 KHz
Symbols for TTI	14
PDCCH & PCFICH (control ch.)	3 symbols
PDSCH (data ch.)	11 symbols
CQI reporting	periodic wideband
CQI processing time	2 TTIs
CQI transmission delay	4 TTIs

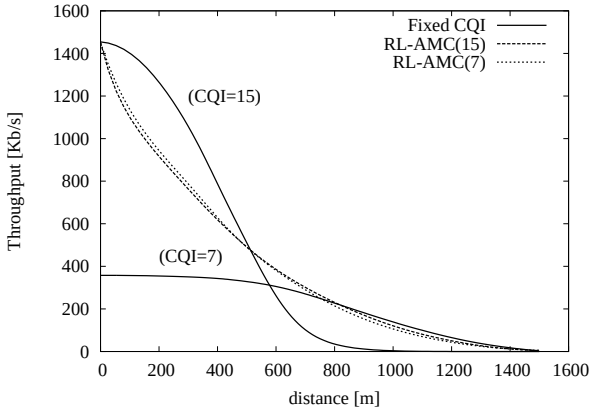

Fig. 2. Average throughput as a function of the distance of the tagged user from the eNB in a pedestrian scenario.

Confidence intervals are very tight and are not shown in the figures. Each simulation run lasts 150 seconds.

B. Results for fixed CQI

In this first set of simulations we assume that ten UEs are randomly deployed in the cell and they are static. Then an additional tagged user is moving with pedestrian speed from the center of the cell to its boundaries. However, independently of the UE position the CQI feedback is constant. Then, Figure 3 shows a comparison of the throughput achieved by the tagged user with and without reinforcement learning. This is obviously a limiting case which is analysed to assess the robustness of our RL-AMC scheme even when CQI provides a very poor prediction of channel performance. As expected with fixed MCS the user throughput is constant when the MCS is over provisioned, while it rapidly goes to zero after a critical distance. On the contrary, our RL-AMC is able to discover the correction factor that should be applied to the initial CQI to force the selection of a more efficient MCS. In addition, the performance of RL-AMC are almost independent of the initial CQI value. Note that in this case RL-AMC must explore the full range of CQI values and we set k in (4) equal to 15.

C. Results with adaptive CQI

In the following experiments we assume that each UE implements the SINR to CQI mapping described in [25]. First of all

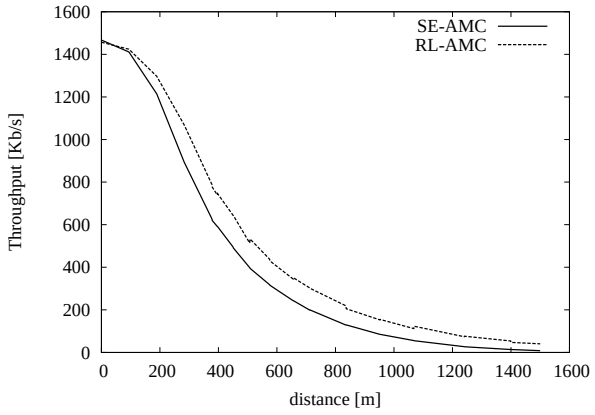

Fig. 3. Average throughput as a function of the distance of the tagged user from the eNB in a pedestrian scenario.

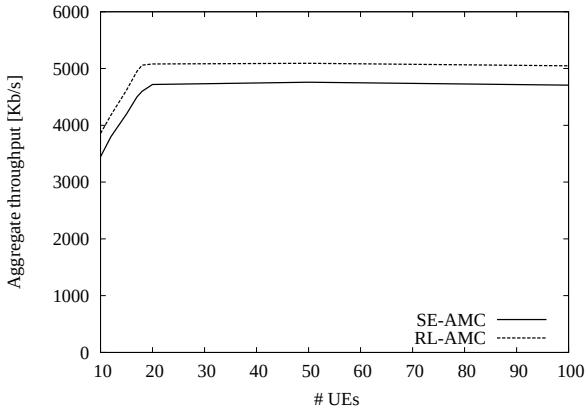

Fig. 4. Average cell throughput as a function of the number of UEs in an urban vehicular scenario.

we consider the same network scenario as in Figure 2, i.e., ten static UEs randomly deployed and one tagged UE moving at pedestrian speed. Then, Figure 3 shows a comparison of the throughput achieved by the tagged user with both SE-AMC and RL-AMC schemes at different distances of the tagged UE from the eNB. We can observe that the MCS selection in SE-AMC is too conservative and this results in a throughput loss. On the contrary, RL-AMC method is able to discover the MCS configuration that can ensure a more efficient use of the available channel resources. This is more evident at intermediate distances from the eNB when short-term fading may lead to use more frequently low-rate MCSs. As shown in the figure, the throughput improvement varies between 20% and 55% in the range of distances between 200 meters and 800 meters.

In the second set of simulations we consider a more dynamic environment in which there is an increasing number of UEs in the cell, and all the UEs are moving according to RWM with speed 30 km/h and pause time equal to 5 seconds. Figure 4 shows a comparison of the aggregate cell throughput with both SE-AMC and RL-AMC schemes as a function of the network congestion (i.e., number of UEs). The results clearly indicate that the throughput improvement provided by RL-AMC is almost independent of the number of UEs and it is about 10%. We can also observe the the cell capacity initially increases when going from 10 to 20 UEs. This is due to two main reasons. First, RR is able to allocate RBs in a more efficient way when the number of UEs is higher. Second, the higher the number of UEs and the higher the probability that one of the UEs is close to the eNB and it can use high data-rate MCSs.

To investigate more in depth the behaviour of the considered AMC schemes, in Figure 5 we show the probability mass function of the number of retransmissions that are needed to successfully transmit a segment of data in a cell with 50 UEs moving as described above. We remind that the same MCS is used for both the first transmission attempt and the eventual subsequent retransmissions. We can observe that with

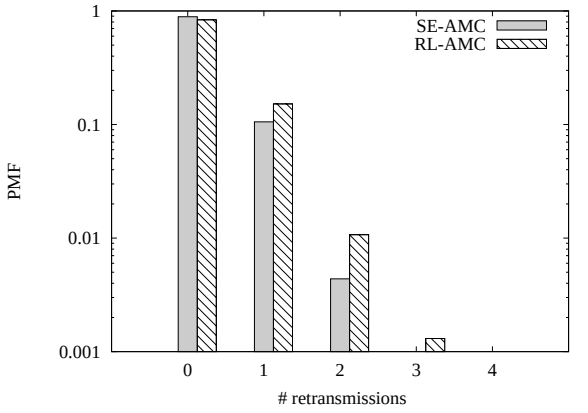

Fig. 5. Probability mass function of the number of retransmissions in an urban vehicular scenario with 50 UEs.

RL-AMC the probability to successfully transmit a segment of data at the first transmission attempt is slightly lower than with SE-AMC. However, the probability of successfully transmiting a segment of data after one or two retransmissions is higher with RL-AMC than with SE-AMC. This confirms our previous observation that the initial MCS selection of SE-MAC is more conservative. On the contrary, RL-AMC is able to also explore MCS with higher data rates when the channel conditions are more favourable and this is beneficial for the throughput performance. Note that this is achieved without violating the BLER requirements imposed by the LTE standard.

VI. Conclusions

In this paper,we have presented a new AMC method for LTE networks that is based on reinforcement learning techniques, We have discussed how inaccurate feedbacks on channel qualities and the complexity of modelling link performance under realistic channel models may easily lead to inaccurate MCS selections. By exploiting reinforcement learning, we can significantly reduce the impact of channel prediction errors on the performance of link adaptation. As future work we plan to explore the use of SINR measurements for directly guiding

the MCS selection. In this case scale-spacing method have to be designed to reduce the state space. A critical extension of this work concerns the investigation of methods to reduce the (typically long) convergence delays of reinforcement learning. To this end recent advancements in RL theory, such as actor-critic methods, will be considered.

REFERENCES

[1] A. Ghosh, R. Ratasuk, B. Mondal, N. Mangalvedhe, and T. Thomas, "LTE-advanced: next-generation wireless broadband technology [Invited Paper]," *IEEE Wireless Communications*, vol. 17, no. 3, pp. 10–22, June 2010.

[2] R. Fantacci, D. Marabissi, D. Tarchi, and I. Habib, "Adaptive modulation and coding techniques for OFDMA systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 9, pp. 4876–4883, Semptember 2009.

[3] 3GPP: Technical Specification Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures (Release 11)," 3GPP TS 36.213 V11.3.0, June 2013.

[4] J. Fan, Q. Yin, G. Li, B. Peng, and X. Zhu, "MCS Selection for Throughput Improvement in Downlink LTE Systems," in *PRoc. of IEEE ICCCN'11*, 2011, pp. 1–5.

[5] J. Francis and N. Mehta, "EESM-Based Link Adaptation in Point-to-Point and Multi-Cell OFDM Systems: Modeling and Analysis," *IEEE Transactions on Wireless Communications*, vol. 13, no. 1, pp. 407–417, January 2014.

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

[7] J. Olmos, S. Ruiz, M. García-Lozano, and D. Martín-Sacristán, "Link Abstraction Models Based on Mutual Information for LTE Downlink," COST 2100, Tech. Rep. 11052, June 2010.

[8] Y. Blankenship, P. Sartori, B. Classon, V. Desai, and K. Baum, "Link error prediction methods for multicarrier systems," in *Proc. of IEE VTC-Fall'04*, vol. 6, 2004, pp. 4175–4179.

[9] K. Brueninghaus, D. Astely, T. Salzer, S. Visuri, A. Alexiou, S. Karger, and G.-A. Seraji, "Link performance models for system level simulations of broadband radio access systems," in *Proc. of IEEE PIMRC'05*, vol. 4, 2005, pp. 2306–2311.

[10] M. Ni, X. Xu, and R. Mathar, "A channel feedback model with robust SINR prediction for LTE systems," in *Proc. of EuCAP'13*, 2013, pp. 1866–1870.

[11] J. Ikuno, S. Pendl, M. Simko, and M. Rupp, "Accurate SINR estimation model for system level simulation of LTE networks," in *Proc. of IEEE ICC'12*, 2012, pp. 1471–1475.

[12] A. Kuhne and A. Klein, "Throughput analysis of multi-user ofdma-systems using imperfect cqi feedback and diversity techniques," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1440–1450, October 2008.

[13] R. Akl, S. Valentin, G. Wunder, and S. Stanczak, "Compensating for CQI Aging By Channel Prediction: The LTE Downlink," in *Proc. of IEEE GLOBECOM'12*, 2012, pp. 4821–4827.

[14] G. Xu and Y. Lu, "Channel and Modulation Selection Based on Support Vector Machines for Cognitive Radio," in *Proc. of WiCOM'06*, 2006, pp. 1–4.

[15] R. Daniels, C. Caramanis, and R. Heath, "Adaptation in Convolutionally Coded MIMO-OFDM Wireless Systems Through Supervised Learning and SNR Ordering," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 1, pp. 114–126, January 2010.

[16] R. Daniels and R. Heath, "Online adaptive modulation and coding with support vector machines," in *Proc. of EW'10*, 2010, pp. 718–724.

[17] J. Leite, P. H. De Carvalho, and R. Vieira, "A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems," in *Proc. of IEEE WCNC'2012*, 2012, pp. 809–814.

[18] Z. He and F. Zhao, "Performance of HARQ with AMC Schemes in LTE Downlink," in *Proc. of IEEE CMC'10*, vol. 2, 2010, pp. 250–254.

[19] P. Tan, Y. Wu, and S. Sun, "Link adaptation based on adaptive modulation and coding for multiple-antenna ofdm system," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 8, pp. 1599–1606, October 2008.

[20] T. Jensen, S. Kant, J. Wehinger, and B. Fleury, "Fast link adaptation for mimo ofdm," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 8, pp. 3766–3778, October 2010.

[21] Donthi, S.N. and Mehta, N.B., "An Accurate Model for EESM and its Application to Analysis of CQI Feedback Schemes and Scheduling in LTE," *IEEE Transactions on Wireless Communications*, vol. 10, no. 10, pp. 3436–3448, October 2011.

[22] T. Tao and A. Czylwik, "Combined fast link adaptation algorithm in LTE systems," in *Proc. of ICST CHINACOM'11*, 2011, pp. 415–420.

[23] R. Sutton and A. Barto, *Reinforcement Learning: An Introduction*. The MIT Press, March 1998.

[24] C. Watkins and P. Dayan, "Q-Learning," *Machine Learning*, vol. 8, pp. 279–292, 1992.

[25] N. Baldo, M. Miozzo, M. Requena-Esteso, and J. Nin-Guerrero, "An Open Source Product-oriented LTE Network Simulator Based on Ns-3," in *Proc. of ACM MSWiM'11*, 2011, pp. 293–298.

[26] 3GPP: Technical Specification Group Radio Access Network, "Conveying MCS and TB size via PDCCH," TSG-RAN WG1 R1-081483, March 2008.

[27] COST Action 231, "Digital mobile radio future generation systems," Final Report - EUR 18957, 1999.

[28] W. Jakes, *Microwave Mobile Communications*. John Wiley & Sons Inc., 1975.

[29] 3GPP: Technical Specification Group Radio Access Network, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception," 3GPP TS 36.104 V11.7.0, January 2014.

[30] D. Karamshuk, C. boldrini, M. Conti, and A. Passarella, "Human mobility models for opportunistic networks," *IEEE Communications Magazine*, vol. 49, no. 12, pp. 157–165, 2011.