

ARIADNE: AI-RAN Informed Link Adaptation in Digital Twin Network Environments

Maria Tsampazi*, Neagin Neasamoni Santhi*, Nicole Perrotta*, Falko Dressler[§], Tommaso Melodia*

*Institute for Intelligent Networked Systems, Northeastern University, Boston, MA, U.S.A.

E-mail: {tsampazi.m, neasamonisanthi.n, perrotta.nic, t.melodia}@northeastern.edu

[§]School of Electrical Engineering and Computer Science, TU Berlin, Germany

E-mail: {dressler}@ccs-labs.org

Abstract—Artificial Intelligence (AI)-powered Radio Access Network (RAN) networks have attracted significant attention from both industry and academia. Meanwhile, Digital Twins offer a safe playground for experimenting with AI/Machine Learning (ML)-based solutions for advanced AI-RAN research. By enabling the testing of online algorithms before deployment on the RAN, they reduce costs and safety risks associated with physical field testing. In this article, we propose ARIADNE, an online Reinforcement Learning (RL)-based module that seamlessly integrates with SIONNA and is tasked with performing link adaptation. We explore different design choices and demonstrate how ARIADNE can surpass industry-standard and state-of-the-art methods by achieving up to 11% and 20% improvements in Spectral Efficiency, respectively. Finally, we show that RL learns a Modulation and Coding Scheme (MCS) selection strategy that diverges from Outer Loop Link Adaptation (OLLA), exhibiting either more conservative or more aggressive behavior depending on the configuration, a trend further corroborated by training offline on 5th generation (5G) over-the-air (OTA) measurements.

Index Terms—5G/6G, Adaptive Modulation and Coding, RL, SIONNA

I. INTRODUCTION

Recent years have seen collaboration among leading industry entities such as Nokia and NVIDIA to advance AI-RAN research [1]. AI-native 6G is expected to emerge from simulation, with Digital Twins playing a key role in the train–simulate–deploy–optimize lifecycle [2]. Indeed, Digital Twins are envisioned to enable faster innovation by allowing the evaluation of “what-if” scenarios for future technologies that are not yet available in hardware [3]. By reducing the “concept-to-live” cycle, dense urban environments and complex 6G use cases can be simulated with greater accuracy, supporting the design and delivery of NextG systems while accelerating the adoption and deployment of advanced AI solutions [2].

In this context, NVIDIA’s SIONNA Digital Twin environment [4] provides an all-in-one platform for wireless research, enabling the execution of system-level simulations over ray-traced channels. In particular, SIONNA-SYS allows the on-the-fly integration of plugin AI/ML solutions, facilitating the testing of key features and capabilities essential to AI-native 6G research. Simulated, controlled environments often provide a practical first step for developing AI/ML solutions before conducting OTA experiments, allowing for algorithm refinement and parameter optimization.

At the same time, a trending topic that has gained widespread attention in the research community is the optimization of the

Next Generation Node Base (gNB)’s Medium Access Control (MAC)-layer scheduler. This involves the optimal allocation of Physical Resource Blocks (PRBs) and selection of the MCS index, as well as the tuning of power control mechanisms (e.g., 3GPP-compliant open and closed-loop power control) that adjust transmit power based on predefined Signal-to-Noise-Ratio (SNR) targets to meet Quality of Service (QoS) requirements for different slices and verticals. Importantly, by jointly optimizing power control and MCS selection, Spectral Efficiency can be significantly improved, demonstrating the benefits of a coordinated approach for enhanced network management.

A. Related Work

Numerous works have focused on RL for MCS selection [5], ranging from offline approaches [6] to contextual Multi-Armed Bandit (MAB) solutions [7], [8]. The authors in [9] introduce GrGym, an RL-based framework for MCS selection in WiFi, which leverages a custom gym environment. In [10], the authors formulate an RL framework where a low discount factor enforces myopic MCS selection based on the current channel state. The industry standard for link adaptation is Outer Loop Link Adaptation (OLLA) [11], and modifications involving RL have been introduced for intelligent on-the-fly adaptation [12]. More recently, the authors in [13] propose a gradient descent approach with a learning rate that self-adapts online through knowledge distillation. However, none of the works mentioned above focuses on the integration of an AI module on the fly within high-fidelity Digital Twins.

B. Contributions

Motivated by the suitability of system-level simulations for digital twin environments, we propose ARIADNE, a framework that leverages NVIDIA’s platform SIONNA to integrate an RL-module for AI-RAN networks, enabling learning-based link adaptation through adaptive MCS selection on channels simulated via SIONNA ray tracing. Our module seamlessly integrates with the SIONNA-SYS platform, enabling direct comparison with state-of-the-art link adaptation algorithms (such as the industry-standard OLLA [11] and NVIDIA’s proposed link adaptation scheme, entitled Self-Adaptive Link Adaptation (SALAD) [13]). In this article, we train and evaluate ARIADNE on a variety of high-fidelity ray-traced channels generated with SIONNA, and we demonstrate that online RL outperforms both OLLA and SALAD in MCS selection in terms of performance, as measured by the achieved Spectral Efficiency. Finally, we evaluate the suitability of RL on offline

5G data collected from an OTA 5G testbed, demonstrating its effectiveness on real-world measurements.

II. ARIADNE: LEARNING LINK ADAPTATION VIA REINFORCEMENT LEARNING

Our RL agent operates at the *Fast Link Adaptation* level, where decisions (i.e., the selection of MCS levels for all User Equipments (UEs)) are enforced on a per-slot basis. This requires precise knowledge of the channel quality. However, Channel Quality Information (CQI) reports are subject to feedback delay [14]. In addition, CQI is typically reported over a wide band, while transmission may occur over a narrower band. As a result, the true Signal to Interference plus Noise Ratio (SINR) is not directly observable and must be estimated online at each slot using only past feedback. Consequently, link adaptation requires jointly inferring the channel quality and selecting an appropriate MCS [13]. Therefore, for the current slot's channel quality, we consider two estimation modes: an *Oracle* mode that establishes an upper bound using SIONNA's simulated environment to calculate the SINR, and a predictor mode for SINR estimation that leverages previous CQI reports.

Markov Decision Process (MDP) Formulation and Temporal Resolution. We formulate the MCS selection problem as an MDP defined by the tuple $(\mathcal{S}, \mathcal{A}, R, P, \beta)$. Critical to our design is a 1 : 1 mapping between the RL environment and the 5G New Radio (NR) Physical (PHY)-layer, where each discrete environment step corresponds exactly to one 5G NR time slot.

A. State Space

At each time slot t , the agent observes a state vector $\mathbf{s}_t \in \mathbb{R}^{(N_{\text{cqi}} + N_{\text{harq}} + 5) \cdot U}$, where U is the number of UEs. The state is composed of per-UE feature vectors:

$$\mathbf{s}_t = [\mathbf{s}_t^{(1)}, \dots, \mathbf{s}_t^{(U)}], \quad (1)$$

where each per-UE component $\mathbf{s}_t^{(u)}$ is given by

$$\mathbf{s}_t^{(u)} = [\tilde{\gamma}_{\text{post},t}^{(u)}, \tilde{\gamma}_{\text{eff},t-k}^{(u)}, h_{t-k}^{(u)}, \tilde{m}_{t-1}^{(u)}, \hat{B}_t^{(u)}, \Delta_{\text{offset},t}^{(u)}, \bar{a}_t^{(u)}], \quad (2)$$

and the respective components are defined as follows:

- $\tilde{\gamma}_{\text{post},t}^{(u)}$: An estimate of the current slot's normalized post-equalization SINR, computed as the mean SINR across allocated resource elements after Linear Minimum Mean Square Error (LMMSE) equalization. In *Oracle* mode, this is derived from the current channel realization; in *predictor* mode, it is estimated from past measurements.
- $\tilde{\gamma}_{\text{eff},t-k}^{(u)}$, $k = 1, \dots, N_{\text{cqi}}$: The N_{cqi} most recent effective normalized SINR values reported by the PHY-layer Abstraction.
- $h_{t-k}^{(u)}$, $k = 1, \dots, N_{\text{harq}}$: A window of N_{harq} prior Hybrid Automatic Repeat reQuest (HARQ) outcomes, where $h \in \{-1, 0, 1\}$ denotes unscheduled, NACK, and ACK, respectively.
- $\tilde{m}_{t-1}^{(u)}$: The normalized previously selected MCS index.
- $\hat{B}_t^{(u)}$: The sliding-window Block Error Rate (BLER) estimate over the last N_{bler} scheduled slots, calculated as $\hat{B}_t^{(u)} = \sum \text{NACK} / \sum \text{Scheduled}$.

- $\Delta_{\text{offset},t}^{(u)}$: A normalized asymmetric feedback accumulator updated by $+0.1$ on ACK and -0.9 on NACK.¹
- $\bar{a}_t^{(u)}$: The running ACK rate over the full episode.¹

B. Action Space

The agent performs link adaptation for the 5G NR Physical Downlink Shared Channel (PDSCH). The action $\mathbf{a}_t \in \{0, 1, \dots, 28\}^U$ is a vector of MCS indices, one per UE, as defined in the 5G NR MCS Table 1 [15]. The mapping is performed directly by the policy network as follows in (3):

$$\mathbf{a}_t = \pi(\mathbf{s}_t) = (m_t^{(1)}, \dots, m_t^{(U)}). \quad (3)$$

C. Reward

We aim to capture the fundamental MCS selection tradeoff. Higher MCS indices result in higher Spectral Efficiency, but simultaneously increase the BLER and, consequently, the probability of decoding failure. Conversely, a conservative MCS lowers the BLER but also reduces the Spectral Efficiency. To target the practical throughput–reliability tradeoff, we maximize a reward that captures the effective Spectral Efficiency, which is directly proportional to system goodput.

Instantaneous Reward. At each environment step (i.e., each 5G NR time slot), the agent selects an MCS index for each UE, and the environment subsequently returns a scalar reward based on the successfully delivered Spectral Efficiency in that slot. The instantaneous reward r_t at slot t is defined as the total achieved Spectral Efficiency across all UEs, as expressed in (4):

$$r_t = \sum_{u=1}^U Q(m_t^{(u)}) \cdot R_c(m_t^{(u)}) \cdot \mathbb{1}\{\mathcal{H}_t^{(u)} = \text{ACK}\}, \quad (4)$$

where $Q(m_t^{(u)})$ and $R_c(m_t^{(u)})$ denote the modulation order and the effective code rate, respectively, associated with the MCS index $m_t^{(u)}$ selected for UE u at slot t . The term $\mathbb{1}\{\cdot\}$ denotes the indicator function, which evaluates to 1 if the HARQ outcome $\mathcal{H}_t^{(u)}$ for the u -th UE is a successful acknowledgment (ACK), and 0 otherwise.

Expected Reward. In the SIONNA-SYS simulator, the HARQ feedback is generated by the PHY Abstraction module as a Bernoulli outcome with success probability $\Pr(\text{ACK}_t^{(u)} = 1 \mid \mathbf{s}_t, a_t) = 1 - \text{BLER}(m_t^{(u)}, \gamma_t^{(u)})$, where the BLER depends on the selected MCS index $m_t^{(u)}$ and the effective SINR $\gamma_t^{(u)}$ computed from the post-equalization SINR at slot t . Given that the nominal Spectral Efficiency is defined as $\text{SE}_{\text{nom},t}^{(u)} \triangleq Q(m_t^{(u)}) \cdot R_c(m_t^{(u)})$ for a chosen MCS, the expected per-slot reward is given in (5):

$$\mathbb{E}[r_t \mid \mathbf{s}_t, a_t] = \sum_{u=1}^U \underbrace{Q(m_t^{(u)}) \cdot R_c(m_t^{(u)})}_{\text{SE}_{\text{nom}}(m_t)} \cdot (1 - \text{BLER}(m_t^{(u)}, \gamma_t^{(u)})). \quad (5)$$

Reward Interpretation. The expectation in (5) evaluates the effective Spectral Efficiency, which is directly proportional to

¹Both $\Delta_{\text{offset},t}^{(u)}$ and $\bar{a}_t^{(u)}$ are derived from HARQ feedback; the former weights NACK outcomes $9\times$ more heavily than ACK, making it more sensitive to link failures than the symmetric ACK rate $\bar{a}_t^{(u)}$.

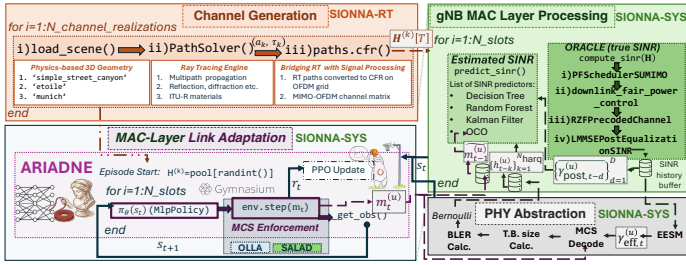


Fig. 1. End-to-end execution of MCS selection in SIONNA. The channel matrix is generated from a pool of ray-traced channel realizations to ensure diversity. The current SINR is either directly used or estimated from previous observations. The post-equalization SINR is then provided as input to the PHY Abstraction. The state s_t is observed by the ARIADNE framework, which selects the corresponding MCS index.

the system goodput. Our RL agent is tasked with maximizing the immediate per-slot reward, which over an episode is equivalent to maximizing the time-average effective throughput. Therefore, our reward formulation naturally captures the fundamental link adaptation tradeoff, with the RL agent learning a policy that selects the MCS to maximize the product $SE_{\text{nom}} \cdot (1 - \text{BLER})$, thereby learning the rate–reliability tradeoff through trial and error.

Adaptive BLER Penalty. Although the reward in (4) implicitly penalizes BLER through zero Spectral Efficiency on NACK, this signal alone may not be sufficient to prevent persistently *aggressive* MCS selection. Link adaptation mechanisms such as OLLA target a fixed BLER level (typically around 10% [11], [13], [15]), balancing throughput and reliability. To incorporate a similar notion of reliability control, we additionally augment the reward with a per-UE penalty weight $\lambda_t^{(u)}$, governed by an integral controller that increases when the cumulative BLER exceeds a target τ and remains zero otherwise, as defined in (6)–(7):

$$r_t = \sum_u \left[SE_{\text{nom},t}^{(u)} \mathbb{1}\{\mathcal{H}_t^{(u)} = \text{ACK}\} - \lambda_t^{(u)} \mathbb{1}\{\mathcal{H}_t^{(u)} = \text{NACK}\} \right], \quad (6)$$

where

$$\lambda_t^{(u)} = \max \left[0, k_E \sum_{i \leq t} \left(\mathbb{1}\{\mathcal{H}_i^{(u)} = \text{NACK}\} - \tau \right) \right]. \quad (7)$$

When $\lambda_t^{(u)} = 0$ for all $u \in \{1, \dots, U\}$, (6) reduces to (4).

D. Transition Dynamics

The environment transitions $P(s_{t+1}|s_t, \mathbf{a}_t)$ are determined by the SIONNA-SYS simulator as detailed in Section III.

III. SYSTEM MODEL

We compare three link adaptation strategies, namely OLLA, SALAD, and ARIADNE. All operate within the same SIONNA-SYS pipeline, and the decision-making component resides within the gNB’s MAC-layer scheduler. Independent of the specific algorithm employed, this reflects a centralized architecture in which the gNB collects CQI and HARQ feedback from all UEs, executes the link adaptation algorithm internally, and assigns MCS indices to scheduled UEs for the upcoming transmission. The per-slot pipeline is illustrated in Fig. 2.

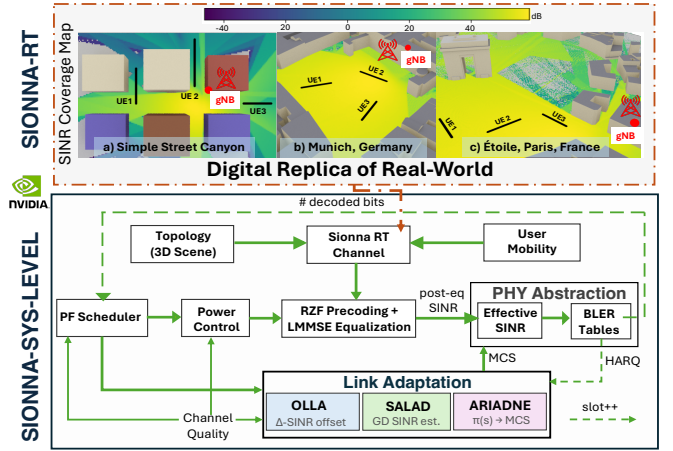


Fig. 2. System-level simulations using SIONNA over ray-traced channels for link adaptation with OLLA, ARIADNE, and SALAD.

A. System Architecture and Link Adaptation Workflow

The end-to-end pipeline is shown in Fig. 1. At the beginning of each simulation, a 3D scene (e.g., Urban Street Canyon, Munich, or Étoile in Fig. 2) is loaded into SIONNA’s ray tracer to compute the channel frequency response \mathbf{H}_t per slot. At each slot, the simulator executes proportional-fair scheduling, power control, and Regularized Zero-Forcing (RZF) precoding with LMMSE equalization, yielding post-equalization SINR values. Link adaptation then selects an MCS index based on the SINR and HARQ feedback. This is the only component that differs across methods (Fig. 2). OLLA maintains a per-UE SINR offset updated via ACK/NACK feedback and selects the highest MCS under a target BLER, using the latest SINR report and HARQ outcome. SALAD primarily relies on ACK/NACK feedback, estimating the SINR online via recursive updates and adapting the BLER target through a feedback control loop. ARIADNE observes a window of past CQI and HARQ feedback, together with a predicted SINR, and directly maps this state to MCS indices via a trained policy network. Finally, PHY Abstraction maps the selected MCS and post-equalization SINR to a BLER, from which the HARQ outcome is sampled, yielding the achieved Spectral Efficiency. All steps except link adaptation are identical across methods, ensuring that any performance difference is solely due to the MCS selection strategy.

B. Implementation and Training

We train the agent using Proximal Policy Optimization (PPO) [16] as implemented in Stable-Baselines3 [17] within a custom Gymnasium environment [18] built on top of the SIONNA system-level simulator, with ARIADNE handling agent–environment interaction. The policy is parameterized by a two-layer Multilayer Perceptron (MLP) with 64 hidden units per layer, a learning rate of 3×10^{-4} , a clip range of $\varepsilon = 0.2$, and an entropy coefficient of 0.05 to encourage exploration, resulting in low computational complexity suitable for gNB deployment. We set the discount factor $\beta = 0$, treating the problem as a *contextual bandit*. This reflects the observation that MCS selection is primarily driven by instantaneous channel state, and temporal credit assignment for future slots does not improve performance. In time-varying fading channels, channel

temporal correlation decreases rapidly, limiting the reliability of long-horizon credit assignment. A myopic policy (i.e., $\beta = 0$) is therefore well suited to this setting. In our setup, we assume three mobile UEs while the ray-traced channels correspond to SIONNA’s Simple Street Canyon scenario.

IV. EXPERIMENTAL EVALUATION

ARIADNE observes a configurable window of N_{cqi} past effective SINR reports and N_{harq} prior HARQ outcomes, as highlighted in Section II-A. We evaluate two configurations, as shown in Table I. In Setup A, the observation window leverages $N_{\text{cqi}} = 3$ reports and $N_{\text{harq}} = 10$ outcomes, providing the agent with temporal context. In Setup B, the observation is limited to a single past SINR report and a single HARQ outcome. Finally, all experimental results reported below have been collected over multiple experiments and channel realizations.

TABLE I

OBSERVATION SIZE CONFIGURATIONS FOR ARIADNE.

Parameter	Setup A	Setup B
CQI window (N_{cqi})	3	1
HARQ window (N_{harq})	10	1

A. Impact of Perfect and Imperfect Channel State Information

We begin our experimental evaluation by quantifying the impact of various SINR estimators on system performance, where the predicted current slot’s normalized post-equalization SINR, denoted as $\tilde{\gamma}_{\text{post},t}^{(u)}$, serves as input to the RL model, as described in Section II-A. Performance is evaluated in terms of the achieved mean *Spectral Efficiency* (SE) and mean BLER. In detail, when using the Oracle as input to ARIADNE’s RL state, we rely on the ground-truth SINR at the current slot, as reported by SIONNA-SYS. When using one of the predictors, namely Decision Tree (DT) [19], Kalman Filter (KF) [20], Random Forest (RF) [21], and Online Convex Optimization (OCO) [14], we replace the Oracle with the predicted SINR in order to assess how closely the resulting performance matches that of the Oracle, which serves as an upper bound. We then compare all the aforementioned RL-based methods against state-of-the-art baselines, namely OLLA and SALAD. Finally, we also discard the current slot’s predicted SINR as input to the RL model (denoted as delayed CQI (dCQI) in the plots) and instead rely solely on past CQI observations, evaluating the resulting performance across all methods.

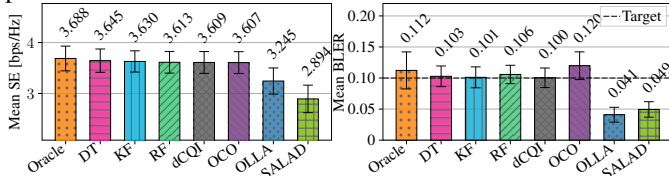


Fig. 3. Mean Spectral Efficiency and BLER for all RL methods under Setup A from Table I.

Regarding the inputs used by each SINR predictor², KF, DT, and RF leverage past SINR observations, KF and OCO incor-

²We do not aim to directly compare the predictors. Instead, we assess whether the performance of the RL agent is affected by the predictor choice, focusing on the robustness of the approach. The KF is a constant-velocity filter with HARQ-driven bias correction, adaptive process noise, and innovation gating. The DT and RF are trained online on past SINR observations, and retrained every N_{slots} . The OCO follows the ACK/NACK-only feedback setting of [14], using mirror descent with Fixed-Share expert mixing over 12 experts defined on the grid $\eta \in \{0.5, 1, 2, 3\}$, $\beta \in \{0, 0.15, 0.3\}$, with $\alpha = 0$.

porate the most recent HARQ feedback, while DT and RF rely on the HARQ accumulator (i.e., $\Delta_{\text{offset},t}^{(u)}$) to capture feedback trends. In contrast, in our implementation, OCO does not rely on explicit SINR history and instead operates using only HARQ outcomes and the previously selected MCS index [14].

TABLE II

PERFORMANCE COMPARISON WITH RESPECT TO THE ORACLE.

Method	SE [bps/Hz]	Δ SE vs. Oracle [%]	BLER
Oracle RL	3.688	0.0	0.112
DT RL	3.645	-1.17	0.103
KF RL	3.630	-1.57	0.101
RF RL	3.613	-2.03	0.106
dCQI RL	3.609	-2.14	0.100
OCO RL	3.607	-2.20	0.120
OLLA	3.245	-12.01	0.041
SALAD	2.894	-21.53	0.049

In Fig. 3, we observe that ARIADNE, regardless of the predictor used, consistently outperforms OLLA and SALAD. Even without prediction of the current slot’s SINR, denoted as dCQI in the plots, the RL agent relying only on past CQI reports outperforms OLLA by 10% in terms of spectral efficiency, achieving a mean spectral efficiency of 3.609 bps/Hz compared to 3.245 bps/Hz, at the cost of a higher mean BLER of 0.1, whereas OLLA achieves a BLER of 0.041. Finally, OLLA outperforms SALAD by approximately 11%, while achieving similar BLER values. For the configuration of SALAD, we use the default parameters as specified in [13].³ It is worth noting that SALAD may require further parameter tuning for the specific scenario at hand, and therefore the parameterization used here may not be optimal for the current setup. This further highlights the advantage of RL in adapting to changing channel conditions and dynamically adjusting its policy.

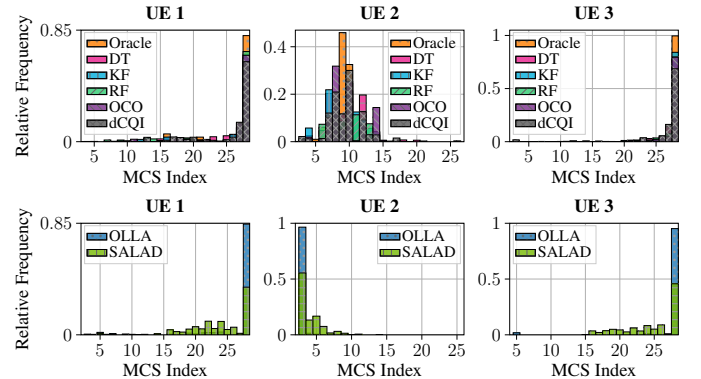


Fig. 4. Distribution of Selected MCS Indices across Methods and UEs under Setup A from Table I.

In Table II, we provide a summary of how closely all methods perform relative to the Oracle when used as input to ARIADNE. We observe that DT achieves a Spectral Efficiency closest to the Oracle, with only $\sim 1\%$ lower performance, reporting a mean Spectral Efficiency of 3.645 bps/Hz compared to 3.688 bps/Hz for the Oracle, followed by KF, whose Spectral Efficiency is reported at 3.630 bps/Hz. All RL methods achieve a Spectral Efficiency close to the Oracle, within $\sim 2\%$. Finally, OLLA reports a mean Spectral Efficiency that is $\sim 12\%$ lower compared

³SALAD hyperparameters follow the reference implementation [22]: learning rate $\varepsilon = 1.2$, bias score threshold $\rho = 0.25$, score window $T = 15$, probing probability $p^{\text{probe}} = 0.15$, probing BLER target $\tau^{\text{probe}} = 0.95$, and integral gain $k_E = 0.05$.

to the Oracle, while SALAD underperforms all methods by $\sim 20\%$. However, all RL methods report a mean BLER of 0.1, indicating the tradeoff between high throughput and reliability. In contrast, OLLA and SALAD maintain the BLER well below the 10% target, at the cost of lower Spectral Efficiency.

Fig. 4 shows the relative frequency distribution of selected MCS indices across different SINR predictor methods and UEs. Notably, the mean SINR observed per UE is ~ 26 dB for UE 1, ~ 6 dB for UE 2, and ~ 28 dB for UE 3. We observe that, for UE 1 and UE 3, all RL methods select the highest MCS indices, indicating agreement across methods in high-SINR regimes. In contrast, for UE 2, which experiences the lowest SINR, the RL agent adopts a more aggressive strategy, selecting MCS indices between 5 and 15, with most selections concentrated around 8–12. On the other hand, OLLA and SALAD predominantly select low MCS indices, concentrated below 10, indicating a more conservative approach.

B. Impact of the Adaptive BLER Penalty

We evaluate the impact of different k_E values, as defined in (6)–(7). Since the previous results do not explicitly constrain the BLER and yield values above 10% for all predictor modes and the Oracle, in contrast to SALAD and OLLA, which maintain BLER below 0.1, we vary $k_E \in \{0, 0.025, 0.1, 0.5\}$. Here, $k_E = 0$ corresponds to no penalty, $k_E = 0.025$ provides a moderate penalty, and $k_E = 0.1$ and $k_E = 0.5$ impose increasingly strong penalties on the median BLER. The resulting performance is summarized in Table III and further illustrated in Fig. 5. As expected, increasing k_E leads to more conservative MCS selection, reducing Spectral Efficiency while improving reliability. In particular, $k_E = 0$ achieves the highest Spectral Efficiency but at the cost of a higher BLER, while larger values of k_E progressively reduce the median BLER at the expense of Spectral Efficiency. Among all configurations, $k_E = 0.1$ provides the best tradeoff, achieving high Spectral Efficiency while consistently maintaining the BLER below the 10% target (Fig. 5). Finally, OLLA and SALAD remain the most conservative.

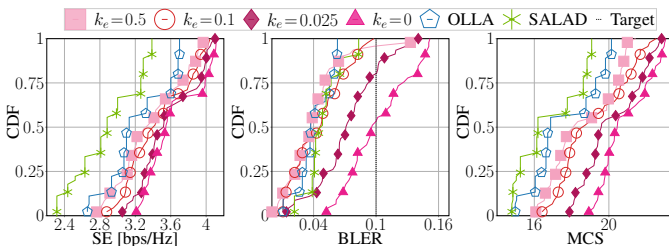


Fig. 5. CDFs of Spectral Efficiency, BLER, and MCS under Setup A from Table I, leveraging the DT predictor.

TABLE III
IMPACT OF k_E ON MEDIAN SPECTRAL EFFICIENCY, BLER, AND MCS

k_E	SE [bps/Hz]	BLER	MCS
0	3.547	0.096	21
0.025	3.520	0.072	20
0.1	3.421	0.047	19
0.5	3.342	0.037	18
OLLA	3.131	0.041	17
SALAD	2.882	0.049	16

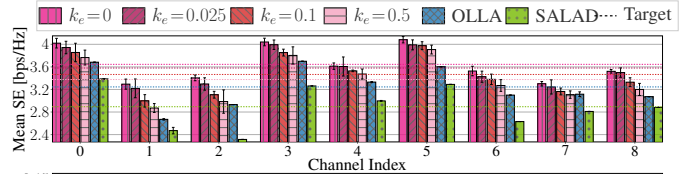


Fig. 6. Spectral Efficiency and BLER under Setup A from Table I, using the DT predictor across multiple channel realizations and experiments.

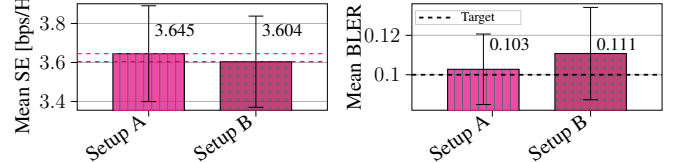


Fig. 7. Mean Spectral Efficiency and BLER for all RL methods and baselines under Setup A and B from Table I, leveraging the DT predictor.

Finally, in Fig. 6, we observe the mean Spectral Efficiency across each channel realization, averaged over multiple experiments. We observe that RL with $k_E = 0.1$ consistently meets the 10% BLER target while achieving one of the highest Spectral Efficiency values among all considered methods, outperforming the industry-standard OLLA. In contrast, RL with $k_E = 0$ and $k_E = 0.025$ achieves higher Spectral Efficiency but often reaches or exceeds the BLER target. Based on these results, we identify $k_E = 0.1$ as the recommended operating point, as it consistently meets the 10% BLER target while achieving high Spectral Efficiency outperforming both OLLA and SALAD.

C. Impact of Observation Window Size

We now proceed by evaluating how different observation window sizes impact the mean performance of the RL agent. Once again, for the current slot's SINR prediction, we leverage the DT, as it achieved the closest performance to the Oracle in terms of Spectral Efficiency in the previous evaluation. The parameterization for this step is given in Table I. As shown in Fig. 7, a smaller observation window (Setup B) results in a slightly lower mean Spectral Efficiency of 3.604 bps/Hz compared to 3.645 bps/Hz achieved by Setup A. Although both Spectral Efficiency values are comparable, Setup A achieves a lower BLER (0.103) than Setup B (0.111). Both values remain above the 10% BLER target, which is expected since the adaptive penalty reward is not leveraged in these experiments. Overall, Setup A is preferred, as it achieves slightly higher Spectral Efficiency while maintaining a lower BLER, enabling the RL agent to select actions without significantly exceeding the BLER target.

V. SITE-SPECIFIC TRAINING AND ROBUSTNESS ACROSS CHANNEL SCENARIOS

Fig. 8 shows the training reward evolution across three different channel scenarios in SIONNA. The RL agent consistently improves its performance and converges in all considered environments. This indicates that the learned policy adapts to varying channel characteristics.

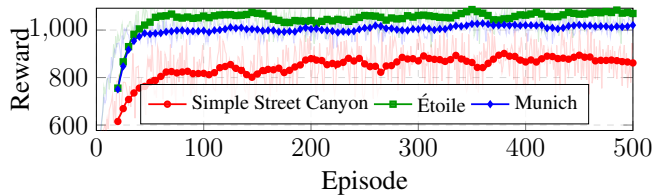


Fig. 8. Training reward over episodes.

VI. FUTURE INTEGRATION OF RL ON OTA 5G TESTBEDS

To examine whether reward-driven MCS selection extends beyond simulation, we train an Fitted Q-Iteration (FQI) agent [23] on OTA measurements collected from a 5G testbed [24], using both Downlink (DL) and Uplink (UL) data. The state consists of CQI, Reference Signal Received Power (RSRP), and instantaneous BLER; the reward is a proxy for achieved throughput derived from logged measurements; and the Q-function is an RF with $\gamma = 0.5$. As shown in Fig. 9, Q-function targets stabilize across Bellman iterations, and the learned policy shifts MCS selections away from the scheduler’s preferred indices. Since these datasets reflect only the OpenAirInterface (OAI) OLLA scheduler’s choices, closed-loop evaluation is left for future work. The fact that both this offline agent and the online PPO agent independently deviate from OLLA suggests that reward-driven MCS selection tends to identify operating points that rule-based schedulers do not.

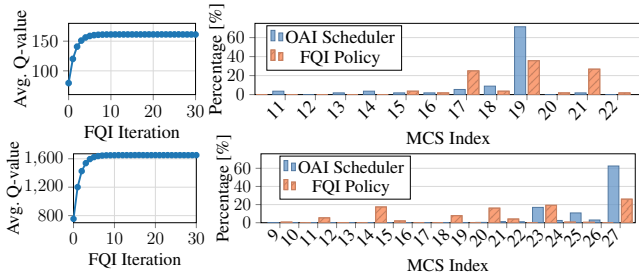


Fig. 9. Offline FQI results on OTA data for UL (top) and DL (bottom). Left: Q-value convergence across Bellman iterations. Right: MCS distribution of the learned policy vs. the OAI scheduler.

VII. CONCLUSIONS AND FUTURE WORK

We developed ARIADNE, a framework that seamlessly integrates with SIONNA, leveraging online RL for MCS selection. The use of system-level simulators, which enable the on-the-fly integration of AI modules, has facilitated the exploration of various approaches, spanning reward formulation and observation window design, to study the impact of different design choices on link adaptation. Although the SIONNA integration provides a practical first step toward the design of RL-based link adaptation, future work will focus on the deployment of ARIADNE on OTA 5G testbeds. Preliminary results on both DL and UL 5G data demonstrate the potential of learning-based approaches in real-world scenarios.

REFERENCES

- [1] NVIDIA Corporation, “NVIDIA and Nokia to Pioneer the AI Platform for 6G — Powering America’s Return to Telecommunications Leadership,” NVIDIA Newsroom, Press Release, Oct. 2025, [Available Online]: <https://nvidianews.nvidia.com/news/nvidia-nokia-ai-telecommunications>.
- [2] Nokia, “Nokia launches Nokia RAN Digital Twin to turbo-charge AI-native 6G, powered by NVIDIA Aerial Omniverse Digital Twin,” Nokia Corporate Blog, Feb. 2026, [Available Online]: <https://www.nokia.com/blog/nokia-launches-nokia-ran-digital-twin-to-turbo-charge-ai-native-6g-powered-by-nvidia-aerial-omniverse-digital-twin/>.
- [3] M. Polese, L. Bonati, S. D’Oro, P. Johari, D. Villa, S. Velumani, R. Gangula, M. Tsampazi, C. P. Robinson, G. Gemmi *et al.*, “Colosseum: The open RAN digital twin,” *IEEE Open Journal of the Communications Society*, 2024.
- [4] J. Hoydis, S. Cammerer, F. Ait Aoudia, M. Nimier-David, A. Keller, A. Vem, M. Stark, and T. O’Shea, “Sionna: An Open-Source Library for Next-Generation Physical Layer Research,” *arXiv preprint arXiv:2203.11854*, 2022.
- [5] N. Khedhri and M. Najar, “Adaptive modulation selection in wireless communications: A comparative study of reinforcement learning, deep learning, deep reinforcement learning, and traditional policies,” in *International Wireless Communications and Mobile Computing (IWCMC)*. IEEE, 2025, pp. 1622–1625.
- [6] S. Peri, A. Russo, G. Fodor, and P. Soldati, “Offline reinforcement learning and sequence modeling for downlink link adaptation,” in *International Conference on Machine Learning for Communication and Networking (ICMLCN)*. IEEE, 2025, pp. 1–7.
- [7] S. K. Pulliyakode and S. Kalyani, “Reinforcement learning techniques for outer loop link adaptation in 4G/5G systems,” *arXiv preprint arXiv:1708.00994*, 2017.
- [8] V. Saxena, J. Jaldén, J. E. Gonzalez, M. Bengtsson, H. Tullberg, and I. Stoica, “Contextual multi-armed bandits for link adaptation in cellular networks,” in *Workshop on Network Meets AI & ML*, 2019, pp. 44–49.
- [9] A. Zubow, S. Rösler, P. Gawlowicz, and F. Dressler, “GrGym: When GNU radio goes to (AI) gym,” in *Proc. of the 22nd International Workshop on Mobile Computing Systems and Applications*, 2021, pp. 8–14.
- [10] J. P. Leite, P. H. P. de Carvalho, and R. D. Vieira, “A flexible framework based on reinforcement learning for adaptive modulation and coding in OFDM wireless systems,” in *Wireless Communications and Networking Conference (WCNC)*. IEEE, 2012, pp. 809–814.
- [11] K. I. Pedersen, G. Monghal, I. Z. Kovacs, T. E. Kolding, A. Pokhariyal, F. Frederiksen, and P. Mogensen, “Frequency domain scheduling for OFDMA with limited and noisy channel feedback,” in *IEEE 66th Vehicular Technology Conference*, 2007, pp. 1792–1796.
- [12] P. Kela, T. Höhne, T. Veijalainen, and H. Abdulrahman, “Reinforcement learning for delay sensitive uplink outer-loop link adaptation,” in *Joint European Conference on Networks and Communications & 6G Summit*. IEEE, 2022, pp. 59–64.
- [13] R. Wiesmayr, L. Maggi, S. Cammerer, J. Hoydis, F. A. Aoudia, and A. Keller, “SALAD: Self-adaptive link adaptation,” *arXiv preprint arXiv:2510.05784*, 2025.
- [14] L. Maggi, B. Bonev, R. Wiesmayr, S. Cammerer, and A. Keller, “SINR Estimation under Limited Feedback via Online Convex Optimization,” *arXiv preprint arXiv:2603.02061*, 2026.
- [15] 3rd Generation Partnership Project, “Nr; physical layer procedures for data (3gpp ts 38.214),” 2023.
- [16] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, “Proximal Policy Optimization Algorithms,” 2017, [Available Online]: <https://arxiv.org/abs/1707.06347>.
- [17] A. Raffin, A. Hill, A. Gleave, A. Kanervisto, M. Ernestus, and N. Dormann, “Stable-Baselines3: Reliable Reinforcement Learning Implementations,” GitHub Repository, 2021, [Available Online]: <https://github.com/DLR-RM/stable-baselines3>.
- [18] Farama Foundation, “Gymnasium: A Standard Interface for Reinforcement Learning Environments,” GitHub Repository, 2023, [Available Online]: <https://github.com/Farama-Foundation/Gymnasium>.
- [19] J. R. Quinlan, “Induction of decision trees,” *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [20] R. E. Kalman, “A New Approach to Linear Filtering and Prediction Problems,” *Transactions of the ASME—Journal of Basic Engineering*, vol. 82, pp. 35–45, 1960.
- [21] L. Breiman, “Random Forests,” *Machine Learning*, vol. 45, 2001.
- [22] NVlabs, “SALAD: Self-Adaptive Link Adaptation — Reference Implementation,” GitHub Repository, NVlabs, 2025, [Available Online]: https://github.com/NVlabs/salad/blob/main/notebooks/meet_salad.ipynb.
- [23] D. Ernst, P. Geurts, and L. Wehenkel, “Tree-based batch mode reinforcement learning,” *Journal of Machine Learning Research*, vol. 6, 2005.
- [24] D. Villa, I. Khan, F. Kaltenberger, N. Hedberg, R. S. da Silva, S. Maxenti, L. Bonati, A. Kelkar, C. Dick, E. Baena *et al.*, “X5G: An open, programmable, multi-vendor, end-to-end, private 5G O-RAN testbed with NVIDIA ARC and OpenAirInterface,” *IEEE Transactions on Mobile Computing*, 2025.