Reinforcement Learning Based Goodput Maximization with Quantized Feedback in URLLC

Hasan Basri Celebi[1,2] and Mikael Skoglund[1]
[1]*KTH Royal Institute of Technology, Stockholm, Sweden*
[2]*Hitachi Energy, Västerås, Sweden*
e-mail: hasan-basri.celebi@hitachienergy.com, skoglund@kth.se

***Abstract*—This paper presents a comprehensive system model for goodput maximization with quantized feedback in Ultra-Reliable Low-Latency Communication (URLLC), focusing on dynamic channel conditions and feedback schemes. The study investigates a communication system, where the receiver provides quantized channel state information to the transmitter. The system adapts its feedback scheme based on reinforcement learning, aiming to maximize goodput while accommodating varying channel statistics. We introduce a novel Rician-K factor estimation technique to enable the communication system to optimize the feedback scheme. This dynamic approach increases the overall performance, making it well-suited for practical URLLC applications where channel statistics vary over time.**

Keywords-URLLC, reinforcement learning, quantized feedback, Rician-K estimation, goodput maximization.

I. Introduction

Ultra-Reliable Low-Latency Communication (URLLC) systems are facing the challenge of achieving reliability with maximum data transmission rates while dynamically responding to fluctuating channel conditions. In this context, goodput, which represents the rate of successful information transmission, is a key metric for evaluating overall system performance. Optimizing goodput emphasizes the significance of feedback mechanisms. These mechanisms let the transmitter adapt its transmission strategies efficiently [1].

A. Related Work

Various research explored scenarios where partial Channel State Information (CSI) is available, aiming to reduce system overhead compared to full feedback approaches. In [2], a more systematic feedback approach is explored, focusing on quantized CSI. Kim et al. [3] investigate wireless communication systems with partial CSI transmitted over an error-free quantized feedback channel in the asymptotic regime, proposing an adaptive feedback scheme to maximize goodput. Recently, advancements in finite blocklength regime for low-latency communication applications have been studied in [4].

On the other hand, significant advancements in URLLC with feedback systems have been highlighted in recent years. [5] addresses URLLC downlink transmission quality challenges ensuring reliability and flexibility in feedback transmission. Authors in [6] introduce Deep-HARQ, an AI-driven algorithm optimizing the interface design for URLLC, significantly reducing link latency. Furthermore, enhancements in downlink link adaptation for URLLC to improve the channel quality while enhancing system-level performance are presented in [7]. These contributions mark significant advancements in reliable URLLC systems with feedback mechanisms [8].

B. Motivation and Contributions

In this study, we assume a quasi-static fading channel where the channel coefficient h remains constant during transmission but varies over different codewords. One of the primary performance metrics in quasi-static fading channels is the system's overall goodput, which can be assessed by the expected rate achieved across a substantial number of packet transmissions with varying transmission rates. This scenario requires a feedback mechanism to transmit the current CSI to the transmitter. Therefore, we propose a system model that investigates the optimum quantized feedback scheme with the purpose of maximizing the overall goodput of the communication system. To extend the existing research, it is also assumed that the channel statistics vary over time due to factors such as mobility and scattering. For this purpose, a Rician distributed channel model is taken into account with an unknown shape factor, which will be defined in the next section.

The foundation of the proposed study lies in a two-part system. The contributions in the current paper can be listed as

- First, we introduce a novel technique for estimating the Rician-K factor, which characterizes the channel's shape factor. This estimate serves as a key input for the second part, where Reinforcement Learning (RL)-based strategies for quantized feedback scheme selection are applied.
- In the second part, we propose a novel RL-based search algorithm to design an adaptive feedback scheme. RL offers a flexible approach to learning and adapting to varying communication environments.

Our approach enables transceivers to dynamically adapt feedback strategies to current channel conditions, thereby maximizing goodput. This model offers a practical solution for dynamic channel adaptation which is crucial for evolving wireless technologies and the increasing demand in next-generation industrial communications for URLLC in the upcoming beyond-5G era.

The remainder of the paper is organized as follows: Section II details the system model and the definition of the problem, considering the varying channel statistics. Section III introduces our novel learning-based Rician-K factor estimator. In Section IV, we present and review the RL-based quantized feedback scheme. Section V provides a performance evaluation of the proposed system, and Section VI concludes the paper.

II. System Model

We consider the discrete-time complex baseband wireless communication system in which the transmitter transmits a codeword over a quasi-static fading channel, where the

complex-valued channel coefficient h is an independent and identically distributed (i.i.d.) random variable according to some distribution but remains constant over the codeword transmission. For the sake of the focus of the study, it is assumed that the receiver has perfect knowledge of h and transmits this information back to the transmitter via an error-free quantized feedback channel.

In such a communication environment, the received signal $\boldsymbol{y}$ can be expressed as

$$\boldsymbol{y} = h\boldsymbol{x} + \boldsymbol{z}, \tag{1}$$

where $\boldsymbol{x}$ and $\boldsymbol{z}$ represent the transmitted codeword and complex Gaussian noise vector where the samples are i.i.d. and $z_i \sim CN(0,1)$, where z_i represents the ith component of $\boldsymbol{z}$.

Let γ denote the i.i.d. channel magnitude which is determined as $\gamma = |h|$, where γ can be defined as a continuous random variable with its corresponding Probability Density Function (PDF), $p(\gamma)$, and Cumulative Distribution Function (CDF), $P(\gamma)$. In this study, it is assumed that both $p(\gamma)$ and $P(\gamma)$ are continuous and $p(\gamma) \geq 0$ over $0 \leq \gamma \leq \infty$.

A. Feedback Channel

It is considered that the receiver divides the positive real line into Λ number of quantization regions and applies a deterministic index mapping on the channel magnitude γ

$$L(\gamma) = l \text{ for } \gamma \in [\lambda_l, \lambda_{l+1}), \tag{2}$$

where $l = 0, 1, \cdots, \Lambda-1$ and $\lambda_0 = 0$ and $\lambda_\Lambda = \infty$. Afterward, the selected index, l, is transmitted back to the transmitter over the error-free feedback channel. Therefore, CSI is partially known to the transmitter.

After receiving the partial CSI, the transmitter selects a transmission rate, described as r_l, which can be defined as the selected rate for the lth quantized region, with the mission of maximizing the goodput of the communication system, which is the maximization of the overall correctly received information rate. For instance, the goodput of a communication system with constant transmission rate r and error rate ϵ is

$$G = r(1-\epsilon). \tag{3}$$

B. Problem Definition

The instantaneous channel capacity, for a given channel magnitude γ with SNR $\mathcal{P}$, is

$$C(\gamma) = \log(1+\gamma^2\mathcal{P}). \tag{4}$$

Suppose $\Lambda = \infty$, which represents perfect CSI at the transmitter, the maximum achievable goodput is the ergodic capacity since it is possible to match the transmission rate to $C(\gamma)$. Thus,

$$G_{\Lambda=\infty} = \int_0^\infty p(\gamma)C(\gamma)\mathrm{d}\gamma. \tag{5}$$

On the other hand, if $\Lambda = 1$, which means no CSI at the transmitter, the maximum achievable goodput can be found by solving the following optimization problem

$$G_{\Lambda=1} = \max_{r\geq 0} \int_{\sqrt{\frac{2^r-1}{\mathcal{P}}}}^\infty r p(\gamma)\mathrm{d}\gamma. \tag{6}$$

When $\Lambda \in [2, \infty)$, determining the maximum achievable goodput value becomes a challenging task. Consequently, the objective becomes identifying the optimal λ_l and r_l configurations that maximize the long-term goodput.

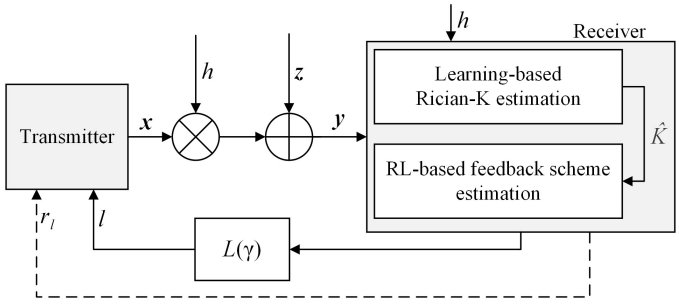

Figure 1. The proposed system model.

Suppose γ_l^r is the reconstruction point of the lth quantization region and the transmitter selects $r_l^r = C(\gamma_l^r)$ as the transmission rate. For a given channel realization γ, where $\lambda_l \leq \gamma < \lambda_{l+1}$, we know that as long as

$$r_l^r \leq \log(1+\gamma^2\mathcal{P}) = C(\gamma), \tag{7}$$

error-free transmission is possible. Otherwise, the communication is in outage. Thus, once the γ_l^rs and r_l^rs for $l = 1, 2, \cdots, \Lambda$ are estimated, the overall outage probability of such a communication system is

$$\sum_{l=0}^{\Lambda-1} (P(\gamma_l^r) - P(\lambda_l)). \tag{8}$$

Therefore, the optimum selections of λ_ls and r_ls can be found by solving the following optimization problem

$$\max_{r_l^r, \lambda_l, \gamma_l^r} \quad \sum_{l=0}^{\Lambda-1} r_l^r \left(P(\lambda_{l+1}) - P(\gamma_l^r)\right) \tag{9a}$$

$$\text{s.t.} \quad 0 < \lambda_l \leq \gamma_l^r < \lambda_{l+1} < \infty. \tag{9b}$$

This problem has been studied in [3], [4], and [9], and it was shown that optimum r_l^rs can be achieved by setting $r_l^r = C(\lambda_l)$ and quantization levels, λ_l^r for $l = 1, 2, \cdots, \Lambda-1$, can be found by solving the following equation with an iterative algorithm

$$P(\lambda_{l+1}^r) = P(\lambda_l^r) + \left(\frac{1}{\mathcal{P}} + \lambda_l^r\right) p(\lambda_l^r) \log\left(\frac{1+\lambda_l^r\mathcal{P}}{1+\lambda_{l-1}^r\mathcal{P}}\right). \tag{10}$$

C. Varying Channel Statistics

Notice that the channel statistics are assumed to be fixed in the optimization problem above. In this paper, we extend these results by *letting the channel statistics vary over time*.

It is assumed that $h \sim CN(\mu, \sigma^2)$, with slowly varying μ over time. Since $h \sim CN(\mu, \sigma^2)$, the channel magnitude, γ, is Rician distributed random variable with varying K-factor, which represents the shape parameter of the distribution and can be defined as the power ratio of the line-of-sight signal power to the remaining multipath and is expressed as [10]

$$K = \mu^2/\sigma^2. \tag{11}$$

This assumption is broader and more realistic for many real-world applications where the channel statistics change due to mobility, scattering, etc. In our analyses, we assumed that μ remains constant for 200 channel realizations and then changes to a new value.

For this purpose, we divide the proposed method into two parts: i) Rician-K factor estimation and ii) RL-based quantized feedback scheme selection. In the first part, we introduce a novel technique for estimating the Rician-K factor. The estimate obtained in this initial phase serves as an input for the subsequent part, where an RL-based strategy is employed to dynamically select and update the feedback scheme to optimize the overall goodput of the communication system, aligning it with the current channel statistics. The proposed system model is shown in Fig. 1.

In Fig. 1, it is worth highlighting the presence of two distinct feedback channels. The first feedback channel, depicted with a solid line, serves as the quantized feedback channel and is utilized in each transmission of a codeword. In contrast, the second feedback channel, indicated by a dashed line, plays a unique role during the training phase, specifically for transmitting updated transmission rates associated with quantization level l. It is important to emphasize that this secondary feedback channel is not employed in every subsequent transmission; rather, its usage is triggered by the decision made by the RL-based feedback scheme to update r_l, as will be discussed in Section IV. In this way, the transmitter's knowledge is restricted to the selected transmission rate r_l for each l, which minimizes the need for additional data transmission during training since all learning algorithms are implemented at the receiver. This also contributes to reducing the overall computational load on transceivers, which is a crucial factor for mitigating the latency due to computational processes [11] and [12].

III. Learning-Based Estimator for Rician-K Factor

Rician-K factor estimation is a well-studied topic in the literature. Moment-based and maximum-likelihood estimators have been presented in [10], [13]. Here, we first introduce the moment-based estimators and then present our findings.

The PDF of the channel magnitude γ is given by

$$p(\gamma) = \frac{2\gamma}{\sigma^2} \exp\Big(\frac{-(\gamma^2+\mu^2)}{\sigma^2}\Big) \mathrm{I}_0\Big(\frac{2\gamma\mu}{\sigma^2}\Big), \tag{12}$$

where $\mathrm{I}_z(\cdot)$ is the modified Bessel function of the first kind with order z. By using the first few raw moments of Rician distribution [14], the following estimators can be found after some straightforward mathematical steps

$$\frac{m_1}{\sqrt{m_2}} = \frac{1}{2}\sqrt{\frac{\pi}{K+1}} L_{\frac{1}{2}}(K), \tag{13}$$

$$\frac{m_4}{m_2^2} = 1 + \frac{(2K+1)}{(K+1)^2}, \tag{14}$$

$$\frac{m_6}{m_2^3} = 6 + \frac{K^2(5K-9)}{(K+1)^3}, \tag{15}$$

where $L_{\frac{1}{2}}(K)$ is the Laguerre polynomial, defined as

$$L_{\frac{1}{2}}(K) = \exp\left(-\frac{K}{2}\right)\left((K+1)\mathrm{I}_0\left(\frac{K}{2}\right) + K\mathrm{I}_1\left(\frac{K}{2}\right)\right), \tag{16}$$

and m_i represents the ith raw moment. By using the estimators above, an estimate of K is obtained numerically by computing the empirical moments and solving the nonlinear equations presented, where a close approximation for the Laguerre polynomial can also be used as in [15].

In this study, we extend these results and propose a novel moment-based learning model for Rician-K estimation. In this proposed method, we employ a more comprehensive set of features unlike the already presented studies in the literature [16], which uses the traditional amplitude samples directly. Specifically, we extract the first ten empirical moments of the Rice-distributed random variable as input features. These moments include the empirical mean, variance, skewness, and kurtosis.

The idea behind using these moments as input features is because of their ability to capture the underlying statistical characteristics of the Rician random variable γ. This approach offers a more detailed and informative representation of the channel compared to conventional moment-based estimators, which often rely only on a few moments. Our learning-based approach is based on the eXtreme-Gradient-Boosting (XGBoost) regression model, which has gained popularity in various domains for its capability to handle complex relationships with high-performance predictions [17]. Other learning algorithms such as linear regression, histogram-based gradient boosting regression, random forest regressor, cat-boost regressor, etc. are also investigated. We skipped their results since their overall performance was worse than XGBoost.

A. Pre-processing

Notice that the number of inputs of the proposed learning-based method does not change with the number of samples collected, which makes the proposed method easily scalable. On the other hand, identifying the number of samples, N, while computing the empirical raw moments becomes a significant design problem. To find the best selection, we have tested various N values and saw that the best performance is obtained when the training dataset is comprised of $N = 100$. However, similar performance results can be achieved with selections of $N > 50$ since XGBoost uses a learning rate to control the model's parameter updates during training [18], which aims to prevent overfitting while maintaining a low bias and ensuring that the model generalizes well to unseen data.

B. Performance Comparison

To test the performance of the proposed method, we resorted to Monte Carlo simulations and compared the results. For this purpose, a training dataset comprising 10^5 Rician-K factors is created. This dataset contains a range of K values, limited within $0 \leq K \leq 100$, and was appended with their respective empirical raw moments, which were computed from randomly generated samples, with each dataset comprising $N = 100$ samples. Even though we set a constant N for the training dataset, we test the performance of the learning-based estimator with various N values, which are $N = \{25, 50, 10^2, 10^3\}$. On the other hand, we select the estimator formulated in (13) as the moment-based estimator due to its leading performance against (14) and (15) since it uses the advantage of the lower order moments [10].

A comparison of the results is presented in Fig. 2, where the sample mean of the predicted K values, $\hat{K}$, are depicted with upper and lower limits of the confidence region, which is

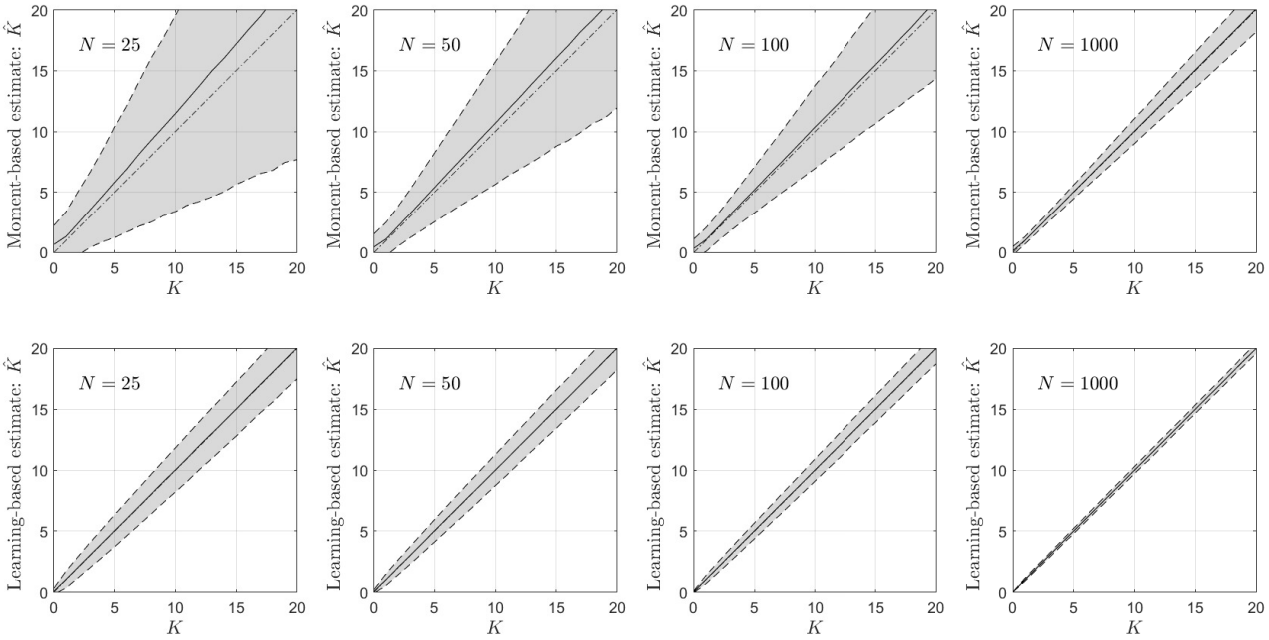

Figure 2. Sample mean and sample confidence region of the two estimators, namely moment-based and learning-based estimators. Results have been depicted for $N = \{25, 50, 100, 1000\}$. (——) Sample mean. (- - -) Upper and lower limits of the confidence region. (—·—) Reference line.

calculated as the $\pm 2\times$standard deviation of $\hat{K}$. Although the estimators are capable of detecting higher K values, we limit the horizontal and vertical axes between $[0 - 20]$ to have a better look at the differences between the estimators.

The key observation derived from Fig. 2 is the remarkable superiority of the learning-based estimator in comparison to the moment-based estimator across all choices of N. Interestingly, the learning-based estimator exhibits better performance, even at $N = 25$, compared to the moment-based estimator's results at $N = 10^2$. Furthermore, the learning-based estimator achieves nearly perfect estimations when $N = 10^3$.

IV. Reinforcement-Learning Based Quantized Feedback Scheme Selection

In this section, we focus on finding the optimum feedback scheme, such as selecting the best λ_ls and r_ls, to maximize the overall goodput of the system. As mentioned in the previous chapter, we use an RL-based search algorithm since the proposed methods in the literature cannot adapt to variable channel statistics.

Let us first denote the optimum selections as λ_l^* and r_l^*. It is shown in [3] that the optimum goodput is achieved by assigning the $r_l^* = C(\lambda_l^*)$. Thus, one can simplify the optimization problem by omitting the r_l variables. Next is the discritization process of the quantization regions. For this purpose, we define a finite number of values for λ_ls and reformulate the optimization problem in (9) as

$$\max_{\lambda_l, \gamma_l} \quad \sum_{l=1}^{\Lambda-1} r_l \big(P(\lambda_{l+1}) - P(\gamma_l)\big) \tag{17a}$$

$$\text{s.t.} \quad 0 \leq \lambda_l \leq \gamma_l < \lambda_{l+1} < \infty, \tag{17b}$$

$$\lambda_l \in \mathcal{S}, \tag{17c}$$

where $\mathcal{S}$ represents the set of finite number of selections. With this reformulation, the solution becomes a sequential search and can be modeled as a Markov decision process and therefore can be solved with RL [19].

A Markov decision process consists of four elements, such as the environment, state space $\mathcal{S}$, action space $\mathcal{A}$, and reward space Ω. In more detail, at each time step t (or at each iteration) the process is in state $s_t \in \mathcal{S}$ and makes a decision and chooses an action $a_t \in \mathcal{A}$. A reward $\omega_t \in \Omega$ is observed after taking the action. Thus, ω_t is received from the environment and is based on s_t and a_t.

For the reformulated optimization problem in (17), states, actions and rewards are designed as follows:

- *State space* $\mathcal{S}$: We set λ_ls for $l = \{1, 2, \cdots, \Lambda - 1\}$ as the agents of the system, except λ_0 and λ_Λ since their locations are fixed. The set of possible selections of λ defines $\mathcal{S}$, which is a subset of R^+. Therefore, s_t is defined to be a vector consisting of the current locations of λ_ls for $l = \{0, 1, 2, \cdots, \Lambda\}$.
- *Action space* $\mathcal{A}$: We design the RL algorithm in such a way that in every subsequent iteration only one agent, e.g. λ_l, can change its status. Thus,

$$a_t \in \{-1, 0, +1\}, \tag{18}$$

where $-1, 0$, and 1 represent decreasing the index of the agent in $\mathcal{S}$ by one, no change, and increasing the index of the agent in $\mathcal{S}$ by one, respectively. Here, we also apply the ϵ-greedy strategy [20], which can be defined as selecting the best action with probability $(1 - \epsilon_t)$ and a uniformly distributed random action with probability ϵ_t, which gives the search algorithm the possibility to explore the whole state space $\mathcal{S}$ without getting stuck into a local maximum. Additionally, it is important to note that any selected action a_t shall not cause any contradiction with the constraint defined in (17b).

- *Reward space* Ω: The reward function is defined as the empirical mean of the goodput that is achieved with the new state s_t, i.e.

$$\omega_t = \frac{1}{M}\sum_{m=1}^{M} C(\lambda_{L(\gamma_m)}) \triangleq G_{\text{emp}}^{M}, \tag{19}$$

where M represents the number of transmitted blocks with state s_t and yet is another design configuration of the RL model which will be discussed in the next section.

A. Q-Learning for Enhancing Goodput

The traditional approach in reinforcement learning involves a method known as Temporal Difference (TD) learning. This technique blends aspects of both Monte Carlo and dynamic programming. It is similar to Monte Carlo since TD learning acquires samples directly from the environment, and it is similar to dynamic programming since it refines its estimates based on both the current and previous assessments. One of the main TD learning methods is Q-learning which can be represented as an RL methodology allowing the agent to acquire the best strategy for navigating a specific environment [21]. This requires the agent to keep track of an approximation of the anticipated long-term discounted rewards for every possible state-action combination and then make choices to maximize these rewards.

In the Q-learning process, the agent iteratively updates its Q-table, which stores the expected cumulative rewards for each state-action pair. The optimal policy can be found by Bellman's optimality equation [21]

$$Q^*(s,a) = \mathrm{E}[\omega_{t+1} + \eta \max_{a'} Q^*(s_{t+1}, a') | s_t = s, a_t = a] \tag{20}$$

where η represents the discount factor which is required to bound the cumulative reward and $\max_{a'} Q^*(s_{t+1}, a')$ defines the best estimate for the next state s_{t+1}. (20) reveals that Q-learning requires the current state-action pair, the resulting reward, and the subsequent state. Thus, the updating process of the Q-table can be formulated as

$$Q(s_t, a_t) = (1-\alpha)Q(s_t, a_t) + \alpha(\omega_t + \eta \max_{a'} Q^*(s_{t+1}, a')), \tag{21}$$

where α is the learning rate.

The total number of possible s_ts in the current problem can be expressed as

$$|\mathcal{S}|!/((\Lambda+1)!(|\mathcal{S}| - \Lambda - 1)!), \tag{22}$$

where $|\cdot|$ represents the cardinality of the set. Given the potential for rapid growth of this number, constructing a comprehensive Q-table covering all possible s_t states becomes impractical. To address this challenge, we adopt a strategy where the algorithm treats each agent independently, thus it only needs the creation of distinct Q-tables for each agent. Additionally, given the variability in channel characteristics, the Q-table must encompass all feasible K values.

Consequently, the Q-learning table size for each agent is determined as $|\mathcal{K}||\mathcal{S}|$, where $\mathcal{K}$ is the set of all possible Rician-K factors. Notably, it remains possible to append these distinct Q-tables into a unified table with size

$$(\Lambda - 1)|\mathcal{K}||\mathcal{S}| + 1. \tag{23}$$

B. Algorithm Design

The proposed algorithm is implemented as follows

- Set M, α, η, and initialize the λ_l values for all agents.
- Initialize the Q-learning table with zeros. Set the ϵ values for all agents, where $\epsilon_1 = 0.5$ in our implementation.
- Start the loop by selecting an agent, i.e. λ_l, where at each iteration a different agent is selected.
- Select an action based on the ϵ-greedy algorithm.
- Update ϵ_t with respect to t, such that $\epsilon_{t+1} = \epsilon_t/\sqrt{t}$.
- After selecting a_t, update λ_l and s_t.
- Send the new r_l to the transmitter.
- Observe M number of transmissions, compute the reward ω_t according to (19), and update the Q-table using (21).

It is important to note that the proposed algorithm may not yield the optimal feedback scheme, which is due to the simplification of the Q-table size, as explained earlier, but it achieves a close approximation.

V. Performance Evaluation

Here, we first show the effect of M on the algorithm performance. For this purpose, we implement a Monte Carlo simulation where $\Lambda = 4$, $K = 10$dB, and $\mathcal{P} = 20$dB for $M = \{10^2, 10^3\}$ and show the average of ω_t at each iteration t with the variance with grey color around the average. Results are depicted in Fig. 3, where the long-term maximum achievable goodput value is also plotted as a dash-dotted line. The results demonstrate that the effect of M is significant for the design of the system. As can be seen from Fig. 3, the algorithm can reach the optimum value in both cases but faster, in terms of t, when $M = 10^3$. However, note that higher M values require more transmission at each iteration. On the other hand, in some cases, ω_t exceeds the upper limit due to its empirical nature.

Next, we focus on the overall performance of the proposed method, where we set $\Lambda = 4$, $\mathcal{P} = 20$dB, $N = 10^2$, and $M = 10^2$, and let K change from 0dB to 10dB then to 20dB. To see the overall performance, we again implement a Monte Carlo environment, from where the average ω_ts at each iteration t are obtained and depicted in Fig. 4. The long-term maximum achievable goodput values for each K are also plotted with dash-dotted lines. It is possible to see that the proposed method can track the change in channel statistics and adapt its feedback scheme so that it can approach the maximum achievable goodput values in every case. It is also important to highlight that, thanks to the RL-based learning approach, once the optimal feedback scheme for a particular Rician-K factor is determined, the transceivers can instantly adjust to the optimal scheme whenever the channel exhibits the same K value.

VI. Conclusions

In this study, we introduce a learning-driven system for goodput maximization with quantized feedback in wireless communication, designed to meet the requirements of URLLC. Our contributions include a novel Rician-K factor estimation technique that improves the adaptability of feedback strategies to changing channel conditions. Additionally, we employed RL to dynamically select and update feedback schemes, demonstrating the system's ability to maximize goodput under evolving channel conditions. The importance of dynamic feedback mechanisms is emphasized, which addresses the unique

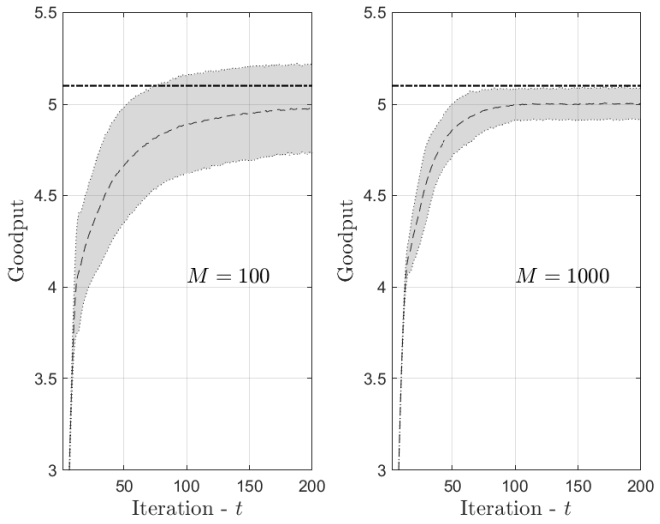

Figure 3. Mean and confidence region of ω_t with respect to iteration t for $M = \{100, 1000\}$ when $\Lambda = 4$, $K = 10$dB, and $\mathcal{P} = 20$dB. (— · —) The long-term average of the maximum achievable goodput. (- - -) Average of ω_t.

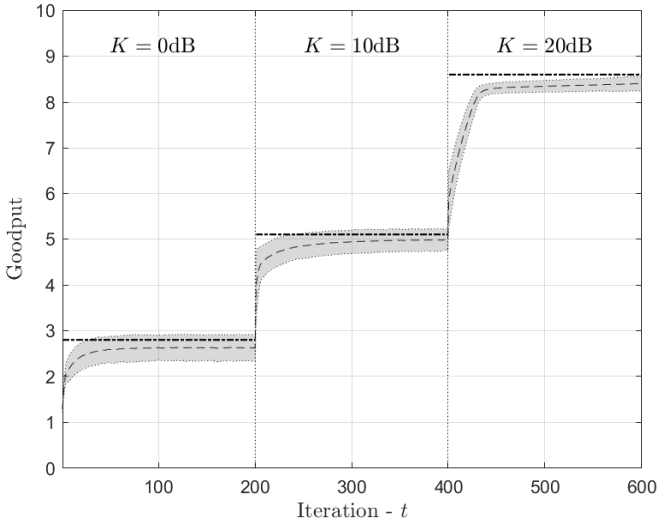

Figure 4. Performance of the proposed method with varying K. (— · —) Long-term average of the maximum achievable goodput. (- - -) Average of ω_t.

challenges posed by URLLC in next-generation wireless networks. Future research could extend the proposed framework to various other wireless communication scenarios.

REFERENCES

[1] H. B. Celebi, "Wireless transmission in future cyber-physical systems," PhD thesis, KTH Royal Institute of Technology, Stockholm, Sweden, 2021.

[2] F. Etemadi and H. Jafarkhani, "Joint source-channel coding for quasi-static fading channels with quantized feedback," in *2007 IEEE International Symposium on Information Theory*, 2007, pp. 2241–2245.

[3] T. T. Kim and M. Skoglund, "On the expected rate of slowly fading channels with quantized side information," *IEEE Transactions on Communications*, vol. 55, no. 4, pp. 820–829, 2007.

[4] H. B. Celebi and M. Skoglund, "Goodput maximization with quantized feedback in the finite blocklength regime for quasi-static channels," *IEEE Transactions on Communications*, vol. 70, no. 8, pp. 5071–5084, 2022.

[5] T.-K. Le, U. Salim, and F. Kaltenberger, "Feedback enhancements for semi-persistent downlink transmissions in ultra-reliable low-latency communication," in *European Conference on Networks and Communications*, 2020, pp. 286–290.

[6] S. AlMarshed, D. Triantafyllopoulou, and K. Moessner, "Deep learning-based estimator for fast HARQ feedback in URLLC," in *IEEE 32nd Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, 2021, pp. 642–647.

[7] G. Pocovi, A. A. Esswie, and K. I. Pedersen, "Channel quality feedback enhancements for accurate URLLC link adaptation in 5G systems," in *IEEE 91st Vehicular Technology Conference*, 2020, pp. 1–6.

[8] H. B. Celebi, A. Pitarokoilis, and M. Skoglund, "A multi-objective optimization framework for URLLC with decoding complexity constraints," *IEEE Transactions on Wireless Communications*, vol. 21, no. 4, pp. 2786–2798, 2022.

[9] B. Makki and T. Eriksson, "On hybrid ARQ and quantized CSI feedback schemes in quasi-static fading channels," *IEEE Transactions on Communications*, vol. 60, no. 4, pp. 986–997, 2012.

[10] A. Abdi, C. Tepedelenlioglu, M. Kaveh, and G. Giannakis, "On the estimation of the k parameter for the rice fading distribution," *IEEE Communications Letters*, vol. 5, no. 3, pp. 92–94, 2001.

[11] H. B. Celebi, A. Pitarokoilis, and M. Skoglund, "Low-latency communication with computational complexity constraints," in *IEEE International Symposium on Wireless Communication Systems*, 2019, pp. 384–388.

[12] H. B. Celebi, A. Pitarokoilis, and M. Skoglund, "Latency and reliability trade-off with computational complexity constraints: OS decoders and generalizations," *IEEE Transactions on Communications*, vol. 69, no. 4, pp. 2080–2092, 2021.

[13] J.-M. Nicolas and F. Tupin, "A new parameterization for the rician distribution," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, pp. 2011–2015, 2020.

[14] S. O. Rice, "Mathematical analysis of random noise," *The Bell System Technical Journal*, vol. 23, no. 3, pp. 282–332, 1944.

[15] H. B. Celebi, A. Pitarokoilis, and M. Skoglund, "Training-assisted channel estimation for low-complexity squared-envelope receivers," in *IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications*, 2018, pp. 1–5.

[16] M. Alymani *et al.*, "Rician k-factor estimation using deep learning," in *29th Wireless and Optical Communications Conference*, 2020, pp. 1–4.

[17] P.-Y. Hsu, I.-W. Yeh, C.-H. Tseng, and S.-J. Lee, "A boosting regression-based method to evaluate the vital essence in semiconductor industry performance," *IEEE Access*, vol. 8, pp. 156 208–156 218, 2020.

[18] D. Nielsen, "Tree boosting with XGBoost," M.S. thesis, Norwegian University of Science and Technology, Trondheim, Norway, 2016.

[19] D. T. Hoang, N. V. Huynh, D. N. Nguyen, E. Hossain, and D. Niyato, "Markov decision process and reinforcement learning," in *Deep Reinforcement Learning for Wireless Communications and Networking*. 2023, pp. 25–36.

[20] A. Masadeh, Z. Wang, and A. E. Kamal, "Reinforcement learning exploration algorithms for energy harvesting communications systems," in *IEEE International Conference on Communications*, 2018, pp. 1–6.

[21] B. Jang, M. Kim, G. Harerimana, and J. W. Kim, "Q-learning algorithms: A comprehensive classification and applications," *IEEE Access*, pp. 133 653–133 667, 2019.