Power and Rate Adaptation for URLLC With Statistical Channel Knowledge and HARQ

Hongsen Peng, Tobias Kallehauge, *Graduate Student Member, IEEE*, Meixia Tao, *Fellow, IEEE*, and Petar Popovski, *Fellow, IEEE*

***Abstract*—This letter investigates a point-to-point ultra-reliable low latency communication (URLLC) transmission with statistical channel knowledge. We consider different hybrid automatic repeat request (HARQ) schemes and investigate the signal-to-noise ratio (SNR) feedback from failed packets to improve transmission efficiency. The problem is formulated as a long-term power minimization problem under URLLC requirement. A deep reinforcement learning (DRL) agent, employing proximal policy optimization (PPO), is used to control transmit power and the coding rate dynamically to solve the formulated problem. Simulation results demonstrate HARQ strategies, SNR feedback, and PPO algorithm can bring significant gains and reveal the impact of reliability and latency.**

***Index Terms*—HARQ, URLLC, deep reinforcement learning.**

I. Introduction

THE 3rd Generation Partnership Project (3GPP) has identified ultra reliable low latency communication (URLLC) as one of the key enablers for 5G wireless networks and beyond. The conflicting requirements of ultra-high reliability and stringent delay constraints inherent to URLLC call for a new class of robust communication strategies [1]. Common methods to reduce the latency includes short packet transmissions, flexible slot structures, and non-coherent transmission without instantaneous channel state information (CSI) at the transmitter [2]. The latency of URLLC can be measured from different communication layers; here we focus on the end-to-end latency at the link layer [1], which takes into account both transmission outage over fading channels and the queueing latency at the transmitter buffer. Under such latency measure, reliability is defined as the probability that the transmitter successfully delivers a finite-size data packet to the receiver within a target time duration.

Designing a communication system that can efficiently meet both the latency and reliability constraints is a significant challenge, especially when considering the fundamental tradeoffs among reliability, delay, throughput, and power consumption. When considering the imperfect CSI, there also exists a tradeoff between pilot length and payload length [3]. It is shown in [4] that single-shot transmission focusing on the physical layer only is quite power-inefficient due to the stochastic wireless channel. This motivates the use of hybrid automatic repeat request (HARQ) schemes to improve the transmission efficiency and potentially bring significant gains by exploiting the previously failed packet through time-diversity [5]. The work [6] analyzed the performance of chase combining HARQ (CC-HARQ) to meet the URLLC reliability and latency constraint while improving overall energy efficiency. In [7], the authors investigated the ability of non-orthogonal multiple access (NOMA) to offer efficient HARQ retransmissions by utilizing the resources concurrently under both statistical CSI and instantaneous CSI assumptions. However, the analysis in [6] and [7] is based on the assumption of constant arrival rate or on-off finite-size data arrival and hence may not be suitable for practical systems with random traffic arrivals.

Another important aspect is to find optimal transmission policies (e.g., coding rate and power) for dynamic URLLC systems. Recently, reinforcement learning (RL) and deep reinforcement learning (DRL) approaches have been utilized for solving the transmission control problem for URLLC [8], [9]. The work [8] proposed an RL-based admission controller that guarantees a probabilistic upper bound on the end-to-end delay of the service system without system model information. This letter, however, assumed that the end-to-end delay is known at any time, which is non-causal. In [9], the authors proposed a DRL algorithm to solve the resource allocation problem with guaranteed reliability and latency. Therein, only the latency at the physical layer is considered without taking into account the queuing delay at the link layer.

In this letter, we study the cross-layer aspects of modeling a point-to-point URLLC link with stochastic traffic arrival process and HARQ retransmission schemes. To our best knowledge, this is the first work that considers stochastic traffic arrival process and HARQ retransmission for the cross-layer design of URLLC systems. The main contributions and results are as follows:

- We establish a novel framework for power and rate adaptation in a URLLC system that employs HARQ with statistical channel knowledge. More specifically, we formulate a long-term power minimization problem subject to delay violation probability and peak power constraints. In addition, we also consider SNR feedback in HARQ and outdated data dropping in queueing model.
- We first transform the formulated problem as an MDP and design an effective reward function to fulfill the delay violation probability constraint. Then we utilize proximal policy optimization (PPO)-based DRL algorithm to learn the transmission control policy (e.g., transmit power and

Manuscript received 14 July 2023; accepted 19 August 2023. Date of publication 30 August 2023; date of current version 11 December 2023. This work was supported in part by the NSF of China under Grant 62125108; in part by the Fundamental Research Funds for the Central Universities of China; and in part by the Villum Investigator Grant "WATER" from the Velux Foundation, Denmark. The associate editor coordinating the review of this article and approving it for publication was T. Q. Duong. *(Corresponding author: Meixia Tao.)*

Hongsen Peng and Meixia Tao are with the Department of Electronic Engineering and the Cooperative Medianet Innovation Center, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: hs-peng@sjtu.edu.cn; mxtao@sjtu.edu.cn).

Tobias Kallehauge and Petar Popovski are with the Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark (e-mail: tkal@es.aau.dk; petarp@es.aau.dk).

Digital Object Identifier 10.1109/LWC.2023.3310205

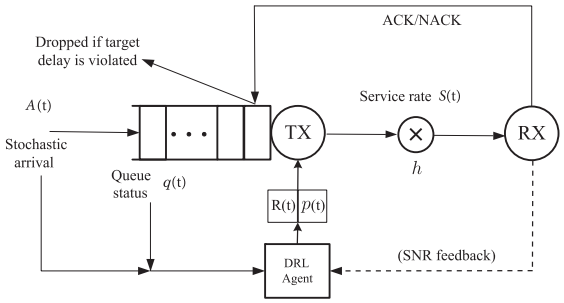

Fig. 1. Transmission model.

coding rate) while ensuring the quality of service (QoS) of URLLC.

Simulation results show that our proposed DRL-based rate and power adaptation policy can significantly reduce the average transmit power compared with the static baseline strategy when simple ARQ is adopted. When HARQ with SNR feedback is considered, the incremental redundancy HARQ (IR-HARQ) mechanism can outperform CC-HARQ significantly due to the more flexible rate and power adaptation while both of them outperform the simple ARQ mechanism.

II. System Model

We consider a point-to-point URLLC transmission over a fading channel as shown in Fig. 1. Data bits arrive at the transmitter (TX) as a stochastic process detailed later. The channel coefficient is assumed to follow Rayleigh fading, i.e., $h \sim \mathcal{CN}(0, \sigma_h^2)$ and keep constant within each time slot consisting of $n \in \mathbb{N}$ channel uses and vary from one slot to another independently. It is also assumed that the instantaneous CSI can be perfectly estimated at the receiver (RX), but only $\sigma_h^2 > 0$ is known at the TX before the transmission. The received signal at the RX is modeled as

$$y = \sqrt{p}hx + z, \tag{1}$$

where p is the transmit power, x is the transmitted signal with unit power, and z is the additive white Gaussian noise (AWGN) with power σ_z^2, i.e., $z \sim \mathcal{CN}(0, \sigma_z^2)$. For simplicity, we assume $\sigma_h^2 = \sigma_z^2 = 1$. The received signal-to-noise ratio (SNR) can thus be expressed as $\gamma = p|h|^2$ which follows exponential distribution with mean p, i.e., $\gamma \sim \exp(p)$.

Due to the lack of instantaneous CSI at the TX, the effect of finite blocklength due to the noise instance is negligible and thus we work with the asymptotic outage probability with infinite blocklength [2]. More specifically, we consider the ε-outage rate, i.e., the maximum rate that the channel can support with probability $1-\varepsilon$ and it is given by:

$$C_\varepsilon = \sup_R \{R : \Pr(\log_2(1+\gamma) < R) < \varepsilon\}. \tag{2}$$

For the Rayleigh block fading, the one-shot ε-outage rate can be derived as

$$C_\varepsilon = \log_2\left[1 - p\ln(1-\varepsilon)\right]. \tag{3}$$

A. Physical Layer Model

In this letter, besides the statistical CSI knowledge at the TX, we also consider the option that the SNRs of the previously failed packets are available at the TX.[1]

[1]In practice, the SNR feedback signal can be combined with ACK/NACK feedback signal and only a few additional bits are needed. Thus the additional time and overhead of the feedback signal can be ignored.

For the scenario without SNR feedback, we employ CC-HARQ. Here, the TX utilizes the same coding scheme, possibly with different transmit powers, for retransmissions. The RX then employs maximal-ratio-combining (MRC) to combine all the received packet replicas effectively increasing the SNR for the transmitted packet. The outage probability using CC-HARQ at the k-th transmission of a packet for a target coding rate R_T is then [10]

$$\mathrm{P}_k^{\mathrm{CC}} = \Pr\left\{\log_2\left(1+\sum_{i=1}^{k}\gamma_i\right) < R_T\right\} \tag{4}$$

$$= \Pr\left\{\sum_{i=1}^{k}\gamma_i < 2^{R_T} - 1\right\}. \tag{5}$$

When the transmit powers $p_1, p_2, \ldots, p_k$ are all distinct, the accumulated SNR, denoted as $Z = \sum_{i=1}^{k}\gamma_i$, follows a hypo-exponential distribution with cumulative distribution function (CDF):

$$F_{\mathrm{HyExp}}(z) = 1 - \sum_{i=1}^{k}\frac{1}{p_i}e^{-\frac{z}{p_i}}\left(\prod_{j=1,j\neq i}^{k}\frac{p_i}{p_i - p_j}\right). \tag{6}$$

When the transmit powers are all the same, i.e., $\forall p_i = p, i \in \{1,2,\ldots,k\}$, the accumulated SNR follows Erlang distribution, whose CDF is

$$F_{\mathrm{Er}}\left(z; k, \frac{1}{p}\right) = \frac{\Upsilon(k, \frac{z}{p})}{\Gamma(k)}, \tag{7}$$

where Υ is the lower incomplete Gamma function and $\Gamma(\cdot)$ is the Gamma function. Here, repeat transmission is employed for simplicity. Thus the outage probability in (5) can be simplified as

$$\mathrm{P}_k^{\mathrm{CC}} = F_{\mathrm{Er}}\left(2^{R_T} - 1; k, \frac{1}{p}\right) = \frac{\Upsilon\left(k, \frac{2^{R_T-1}}{p}\right)}{\Gamma(k)}. \tag{8}$$

For the scenario with SNR feedback, the TX can use failed packets' SNR to further improve the transmission. We consider both CC-HARQ and IR-HARQ. For CC-HARQ with SNR feedback, the outage probability at the k-th transmission changes to a conditional probability given the SNRs of all the previous $k-1$ transmissions. Denoting $\gamma^{(k-1)} = \{\gamma_1, \ldots, \gamma_{k-1}\}$, we have

$$\mathrm{P}_k^{\text{CC-SNR}} = \Pr\left\{\log_2\left(1+\sum_{i=1}^{k-1}\gamma_i + \gamma_k\right) < R_T \,\middle|\, \gamma^{(k-1)}\right\} \tag{9}$$

$$= \Pr\left\{\gamma_k < \Gamma_k^{\mathrm{CC}} \,\middle|\, \gamma^{(k-1)}\right\} \tag{10}$$

$$= 1 - e^{-\frac{\Gamma_k^{\mathrm{CC}}}{p_k}}, \tag{11}$$

where $\Gamma_k^{\mathrm{CC}} \triangleq 2^{R_T} - 1 - \sum_{i=1}^{k-1}\gamma_i$ is known as the residual SNR for CC-HARQ.

For IR-HARQ, the principle is to encode more redundancy bits for each retransmission round, which is then used at the RX to combine and decode the message on a longer codeword [11]. One example of such a scheme uses *rate-compatible punctured convolutional codes* to increase the effective rate of the combined code at each transmission [12].

Hence, instead of increasing the SNR, IR-HARQ increases the code rate for each successive transmission. This gives the following outage probability at the k-th transmission [11]

$$\mathrm{P}_k^{\text{IR-SNR}} = \Pr\left\{\sum_{i=1}^{k}\log_2(1+\gamma_i) < R_T \,\middle|\, \gamma^{(k-1)}\right\} \quad (12)$$

$$= \Pr\left\{\gamma_k < \Gamma_k^{\text{IR}} \,\middle|\, \gamma^{(k-1)}\right\} \quad (13)$$

$$= 1 - e^{-\frac{\Gamma_k^{\text{IR}}}{p_k}}, \quad (14)$$

where $\Gamma_k^{\text{IR}} \triangleq \frac{2^{R_T}}{\prod_{i=1}^{k-1}(1+\gamma_i)} - 1$ is the residual SNR for IR-HARQ.

From both (11) and (14), it is seen that the residual SNR is an important indicator of the uncertainty about the following transmissions, and the larger it is, the higher transmit power is needed in the following retransmission.

B. Link Layer Model

The link layer model is similar to the one used in [13] with the notable innovation of introducing HARQ. The overall procedures are as follows: At the beginning of slot $t\in\mathbb{N}$, the data arrival rate and service rate are denoted as $A(t)$ and $S(t)$, respectively. The arrival process is modeled as a Poisson process with the average arrival rate $\lambda > 0$. The TX initially stores the arrived information bits in a first-in-first-out (FIFO) buffer. When each transmission is over, the RX can reliably detect transmission errors and then send a one-bit error-free and delay-free acknowledgment (ACK) or negative acknowledgment (NACK) signal to inform the TX whether the transmission was successful or not. If the transmission fails, the RX will store the failed packet and wait for retransmission. Simultaneously, upon receiving the NACK signal, the TX will start the next transmission based on the initial transmission. Specifically, when there is no SNR feedback, the TX will adopt CC-HARQ with replica transmission. Otherwise, if the failed packet SNR is available, the TX can use CC-HARQ with the same coding scheme and different transmit power, or IR-HARQ with a different coding rate and different transmit power. The data in the buffer will be removed when an ACK signal is received at the end of each slot, or, when the survival time of the data exceeds the target deadline.

According to the above mechanism, the service rate (in bits per slot) can be expressed as

$$S(t) = \begin{cases} 0, & \text{Outage,} \\ nR_T, & \text{No outage.} \end{cases} \quad (15)$$

At the end of slot t, and before outdated data is dropped, the temporary queue length is given by

$$q_{\text{tmp}}(t+1) = \max\{q(t) + A(t) - S(t), 0\}, \quad (16)$$

where $q(t)$ is the queue length at the beginning of slot t. If the successful transmission has not occurred after $D_{\max}$ slots, the outdated data will be dropped. To characterize this, we first introduce the accumulated number of bits in the latest $D_{\max}$ slots at the end of slot t, i.e.,

$$q_{\text{th}}(t) = \sum_{i=t-D_{\max}+1}^{t} A(i). \quad (17)$$

Then we define the delay experienced by information bits that arrived in slot t as $D(t)$, and the cumulative number of delay violation events can be characterized through queue length violation events as

$$d(t) = \sum_{i=1}^{t}\mathbb{I}\{D(i) > D_{\max}\} = \sum_{i=1}^{t}\mathbb{I}\{q_{\text{tmp}}(i+1) > q_{\text{th}}(i)\}, \quad (18)$$

where $\mathbb{I}\{\cdot\}$ is the indicator function.

With the proactive outdated data dropping (PODD), the queue length $q(t+1)$ at the end of slot t after the outdated data dropping step follows that

$$q(t+1) = \min\{q_{\text{tmp}}(t+1), q_{\text{th}}(t)\}. \quad (19)$$

Finally, we see that the delay violation probability (DVP) is transformed into the queue length violation probability explicitly before outdated data dropping, i.e.,

$$\Pr(D > D_{\max}) = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{I}\{q_{\text{tmp}}(t+1) > q_{\text{th}}(t)\}. \quad (20)$$

C. Problem Formulation

We aim to provide the stochastic arrival process with guaranteed QoS (i.e., bounded DVP) while minimizing the long-term average transmit power. The problem is two-fold. The first one concerns the transmission of a new packet, which is a common part for the different HARQ mechanisms. Here, the aim is to determine the target coding rate R_T and transmit power based on the observed queue length and instantaneous arrived rate. The second problem is about the retransmission of the failed packets, which depends on the HARQ mechanism. Here, the aim is to transmit additional information to the RX with minimum power. Depending on the acquisition of the failed packet SNR, different HARQ mechanisms can be employed as described in Section II-A. Thus we formulate the following long-term power and rate adaptation problem

$$\min_{p(t),R(t)} \quad \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}p(t) \quad (21a)$$

$$\text{s.t.} \quad \Pr(D > D_{\max}) \le \xi, \quad (21b)$$

$$0 \le p(t) \le p_{\max}, \quad (21c)$$

where $p(t)$ and $R(t)$ are the transmit power and the coding rate in slot t, respectively, ξ is the target delay violation probability and $p_{\max}$ is the peak power constraint. When a new transmission starts at slot t, the problem is to determine the target coding rate $R_T = R(t)$ and the transmit power $p(t)$. If the new transmission fails, i.e., when $\log_2(1+p(t)|h(t)|^2) < R_T$, for the subsequent retransmissions, $R(t+i)$ is set according to the physical layer model described in Section II-A. More specifically, $R(t+i) = R(t)$ for CC-HARQ and $R(t+i)$ is the incremental rate for IR-HARQ, where $i \in \{1, 2, 3, \ldots\}$ is the retransmission number. This problem is non-convex, hence it is nontrivial to find the optimal solution. Thus in this letter, we adopt a DRL-based approach.

III. Deep Reinforcement Learning Based Approach

In this section, we first model the power and rate adaptation problem in (21a) as an MDP. Then we will propose a DRL-based approach to solve the MDP.

A. Problem Formulation as an MDP

For simplicity, we assume the MDP is divided into episodes each consisting of consecutive T time slots, where one slot corresponds to one step in the MDP. The elements of MDP for our considered problem are as follows:

State space: The state space $\mathcal{S}$ comprises the queue length, the number of the arrived information bits, delay violations within the current episode, and the number of transmissions of the current packet. When SNR feedback is available, the state space also contains the residual SNR of the kth retransmission Γ_k. When the transmission is successful or the outdated data is removed, the residual SNR is set to zero and the transmission counter is set to 1, which indicates a new packet should be transmitted. Hence the state in the slot t is either given by $s(t) = \{q(t), A(t), d(t), k\}$ (no SNR feedback) or $s(t) = \{q(t), A(t), d(t), k, \Gamma_k\}$ (with SNR feedback).

Action space: The action space $\mathcal{A}$ is composed of the coding rate and the transmit power in the feasible regions with respect to different HARQ mechanisms. Thus the action in slot t can be expressed as $a(t) = \{R(t), p(t)\}$.

B. Reward Function and Algorithm Design

The reward function is critical for learning the optimal transmission policy. It must be designed not only to minimize average transmit power but also to satisfy the requirement of URLLC. We adopt a double-layer penalty reward so as to capture the queue length violation probability (i.e., DVP) as

$$r(t) = \begin{cases} -p(t), & q_{\text{tmp}}(t+1) \le q_{\text{th}}(t) \\ -p(t) - w(t), & q_{\text{tmp}}(t+1) > q_{\text{th}}(t) \end{cases}, \quad (22)$$

$$w(t) = \begin{cases} \Delta\left(\frac{d(t)}{T\xi}\right)^{\beta}, & d(t) \le T\xi \\ \Delta, & d(t) > T\xi \end{cases}, \quad (23)$$

where Δ is a large penalty term, $T\xi$ is the target number of delay violation events which ensures the DVP and β is a positive integer that controls the penalty term $w(t)$ as a function of $d(t)$.

We employ PPO, a state-of-the-art DRL algorithm, to solve the problem due to its simplicity, sample efficiency, stability, and scalability. PPO is a novel policy gradient algorithm derived from the trust region policy optimization (TRPO) algorithm. It can guarantee monotonically non-decreasing performance by optimizing the policy within the divergence constraint (named trust region) [14]. To characterize the difference between the old and new strategies, denote a probability ratio $r_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_n}(a_t|s_t)}$, where π_θ is the new policy and π_{θ_n} is the old policy. The advantage function, which captures how better the current action is than the average policy, is denoted as

$$A_t^{\pi_{\theta_n}} = Q_{\pi_{\theta_n}}(s_t, a_t) - V_{\pi_{\theta_n}}(s_t), \quad (24)$$

where Q is the action value function estimated from transition samples and V is the state-value function. The problem is to improve the new policy by utilizing the advantage functions obtained by the old policy through importance sampling

$$\max_{\theta} \mathop{\mathbb{E}}_{s,a\sim\pi_{\theta_n}} \left[\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_n}(a_t|s_t)} A_t^{\pi_{\theta_n}} \right], \quad (25)$$

with Kullback-Leibler (KL) divergence constraint $\mathbb{E}[D_{\text{KL}}[\pi_\theta||\pi_{\theta_n}]] \le D_{tar}$ between the new and old policies.

Nevertheless, the complicated KL divergence constraint involved in TRPO makes it computationally inefficient and difficult to scale up for large-scale problems. To address this problem, PPO puts the KL divergence constraint into the objective function and adopts a clipping mechanism that allows it to use a first-order optimization and thus can reduce the computational complexity significantly. The optimization problem in PPO is shown in (26), shown at the bottom of the page, and the stochastic gradient ascent algorithm is utilized to update the network. More details can be found in [14].

IV. Simulation Results

A. Simulation Setup

We adopt a similar simulation setting as in [3]. Specifically, each slot contains $n = 200$ channel uses and the delay target is a few slots. To characterize the DVP more precisely, we assume each episode consists of $10/\xi$ steps and there are $400\times10/\xi$ steps. The reward penalty term $\Delta = 20\times p_{\max}$ and $\beta = 16$. $p_{\max}$ is calculated as $p_{\max} = -\frac{2^{\lambda/n}-1}{\log\left[1-(\xi^{D_{\max}^{-1}}/D_{\max})\right]}$ empirically. With respect to PPO, we adopt PPO-clip with a clip ratio $\epsilon = 0.2$. The generalized advantage estimation (GAE) parameter $\lambda_{\text{GAE}} = 0.95$ and the reward discount factor $\gamma_{\text{DIS}} = 0.99$. Both the actor-network and critic-network learning rates are set to $l_r^a = l_r^c = 2\times10^{-4}$. The batch size is 2048 and the minibatch size is 128. The hidden layers of the actor and critic network are fully connected networks with 128 neurons, and the Tanh function is used as activation.

All the results are averaged through 50 times over the trained converged network and all schemes achieve the reliability and power constraints in (21b) and (21c).

B. Numerical Results

We first compare the performance of the following schemes with $\xi = 1\%$ and $D_{\max} = 5$ slots in Fig. 2:

- Effective capacity[2] based static transmission strategy
- Simple ARQ without PODD
- Simple ARQ
- CCARQ
- CCARQ, SNR feedback
- IRARQ, SNR feedback

Firstly, for simple ARQ without PODD, the PPO algorithm is much more efficient than the static effective capacity scheme. As the average arrival rate increases, there exists an almost constant gap between these two schemes. Secondly, it is seen that there is a gap between PODD and non-PODD, which

[2]Effective capacity is derived from large deviation theory and can provide statistical QoS guarantee through static transmission strategy. Due to the page limit, we do not show the details in this letter and similar baseline can be found in our previous work [13].

$$\max_{\theta} \mathop{\mathbb{E}}_{s_t,a_t\sim\pi_{\theta_n}} \left[\min\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_n}(a_t|s_t)}, \text{clip}\left(\frac{\pi_\theta(a_t|s_t)}{\pi_{\theta_n}(a_t|s_t)}, 1-\epsilon, 1+\epsilon \right) \right) A_t^{\pi_{\theta_n}} \right], \ \text{clip}(x,a,b) = \begin{cases} x, & a < x < b \\ a, & x \le a \\ b, & x \ge b \end{cases} \quad (26)$$

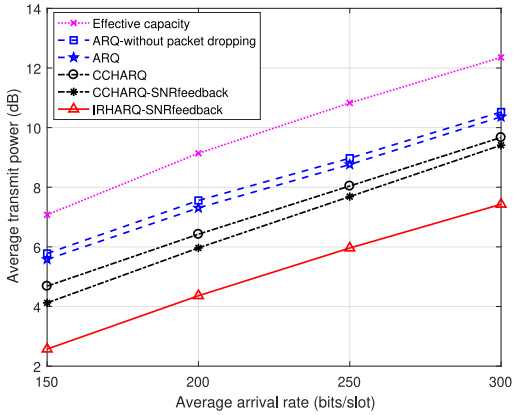

Fig. 2. Performance comparison of different schemes.

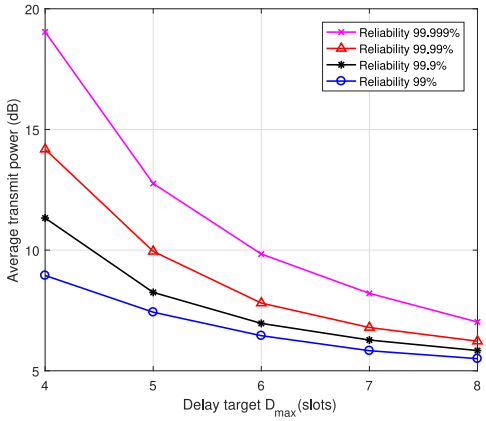

Fig. 3. Impact of reliability and latency.

shows the effectiveness of the PODD mechanism. When CC-HARQ is adopted, the performance is improved compared with simple ARQ with PODD, and the gap is also constant as the average arrival rate increases. With SNR feedback, the performance of CC-HARQ is also slightly improved. This is due to the inflexible repetitive coding scheme of CC-HARQ in the retransmission phase, even if the residual SNR is known. Finally, with SNR feedback and IR-HARQ, the performance is enhanced significantly. This is due to the flexible coding rate and power adaptation during the retransmissions of IR-HARQ.

In Fig. 3, we show the average transmit power as a function of the delay target $D_{\max}$ with different reliability levels $1-\xi$. Here we adopt IR-HARQ with SNR feedback as our transmission scheme and $\lambda = 300$ bits/slot. For each given reliability, we can see that as $D_{\max}$ decreases from 8 to 4, the required transmit power increases and the increasing rate is speeding up. Moreover, as the reliability increases from 99% ($\xi = 10^{-2}$) to 99.999% ($\xi = 10^{-5}$), more transmit power is needed at smaller $D_{\max}$. In particular, when $D_{\max} = 4$ slots, about 10.1 dB more power is required to increase the reliability from 99% to 99.999%, while only 1.5 dB more power is needed if $D_{\max}$= 8 slots. In general, we can conclude that with higher reliability $1-\xi$ and more stringent delay target $D_{\max}$, much more average transmit power is required. This observation coincides with the intuition that, the required transmit power increases more rapidly as the QoS requirements become more stringent.

V. Conclusion

We have investigated the power and rate adaptation problem for URLLC with HARQ with statistical CSI. Specifically, we have employed HARQ transmission scheme and DRL approach to minimize the long-term average transmit power based on the dynamic queueing system, while ensuring the stringent QoS requirements. The simulation results demonstrate that the performance of the proposed PPO-based algorithm significantly outperforms the effective capacity-based approach. It is also shown that PODD, CC-HARQ mechanism, and SNR feedback can also reduce the average transmit power. More importantly, when SNR feedback is available, IR-HARQ outperforms significantly CC-HARQ due to the flexible coding rate and transmit power adaptation. Furthermore, we show the impact of reliability and latency. Future work will consider the adaptation of these strategies to the case of multiple services with heterogeneous timing requirements.

References

[1] P. Popovski et al., "Wireless access in ultra-reliable low-latency communication (URLLC)," *IEEE Trans. Commun.*, vol. 67, no. 8, pp. 5783–5801, Aug. 2019.

[2] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Sep. 2016.

[3] S. Schiessl, H. Al-Zubaidy, M. Skoglund, and J. Gross, "Delay performance of wireless communications with imperfect CSI and finite-length coding," *IEEE Trans. Commun.*, vol. 66, no. 12, pp. 6527–6541, Dec. 2018.

[4] A. Anand and G. de Veciana, "Resource allocation and HARQ optimization for URLLC traffic in 5G wireless networks," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 11, pp. 2411–2421, Nov. 2018.

[5] A. Ahmed, A. Al-Dweik, Y. Iraqi, H. Mukhtar, M. Naeem, and E. Hossain, "Hybrid automatic repeat request (HARQ) in wireless communications systems and standards: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 4, pp. 2711–2752, 4th Quart., 2021.

[6] J. P. Battistella Nadas, O. Onireti, R. D. Souza, H. Alves, G. Brante, and M. A. Imran, "Performance analysis of hybrid ARQ for ultra-reliable low latency communications," *IEEE Sensors J.*, vol. 19, no. 9, pp. 3521–3531, May 2019.

[7] R. Kotaba, C. N. Manchón, T. Balercia, and P. Popovski, "How URLLC can benefit from NOMA-based retransmissions," *IEEE Trans. Wireless Commun.*, vol. 20, no. 3, pp. 1684–1699, Mar. 2021.

[8] M. Raeis, A. Tizghadam, and A. Leon-Garcia, "Reinforcement learning-based admission control in delay-sensitive service systems," in *Proc. IEEE GLOBECOM*, 2020, pp. 1–6.

[9] A. Kasgari, W. Saad, M. Mozaffari, and H. Poor, "Experienced deep reinforcement learning with generative adversarial networks (GANs) for model-free ultra reliable low latency communication," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 884–899, Feb. 2021.

[10] T. V. K. Chaitanya and E. G. Larsson, "Optimal power allocation for hybrid ARQ with chase combining in i.i.d. Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 61, no. 5, pp. 1835–1846, May 2013.

[11] T. V. K. Chaitanya and E. G. Larsson, "Outage-optimal power allocation for hybrid ARQ with incremental redundancy," *IEEE Trans. Wireless Commun.*, vol. 10, no. 7, pp. 2069–2074, Jul. 2011.

[12] J. Hagenauer, "Rate-compatible punctured convolutional codes (RCPC codes) and their applications," *IEEE Trans. Commun.*, vol. 36, no. 4, pp. 389–400, Apr. 1988.

[13] H. Peng, M. Tao, T. Kallehauge, and P. Popovski, "Power adaptation in URLLC over parallel fading channels in the finite Blocklength regime," in *Proc. IEEE GLOBECOM*, Dec, 2022, pp. 1819–1824.

[14] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov. "Proximal policy optimization algorithms." 2017. [Online]. Available: http://arxiv.org/abs/1707.06347