JOINT OPTIMIZATION OF IRS AND UAV-TRAJECTORY

For Supporting Statistical Delay and Error-Rate Bounded QoS Over mURLLC-Driven 6G Mobile Wireless Networks Using FBC

Xi Zhang, Jingqing Wang, and H. Vincent Poor

Massive ultra-reliable and low-latency communications (mURLLC) has been developed as a new and dominating 6G-standard services to support statistical quality of service (QoS) provisioning while raising several major design issues, including massive connectivity, ultra-low latency, and super-reliability. Correspondingly, a number of emerging 6G candidate enablers, including statistical delay and error-rate bounded QoS provisioning, finite blocklength coding (FBC), intelligent reflecting surfaces (IRS), unmanned aerial vehicle (UAV), etc., have been developed to support mURLLC. Specifically, due to the potential improvements in coverage capability as well as the spectral efficiency, both IRS and UAV have been widely proposed to reconfigure wireless propagation environments to compensate for blocked line-of-sight (LOS) communication links and create controllable and smart radio environments. In addition, to solve the massive connectivity issues imposed by mURLLC, integrating UAV with IRS provides a promising means to significantly enhance LOS coverage due to the relatively high altitude and 3D mobility of the UAVs. However, although small-packet communications enabled by FBC are usually employed

Digital Object Identifier 10.1109/MVT.2022.3158047
Date of current version: 24 May 2022

To solve the massive connectivity issues imposed by mURLLC, integrating UAV with IRS provides a promising means to significantly enhance LOS coverage due to the relatively high altitude and 3D mobility of the UAVs.

for massive access to reduce access latency and decoding complexity, how to upper-bound both delay and error rate while efficiently supporting mURLLC in IRS-UAV-integrated systems still remains a challenging problem.

To overcome these difficulties, in this article, we propose FBC-based joint IRS-deployment and UAV-trajectory optimization schemes to support statistical delay and error-rate bounded QoS provisioning for mURLLC over 6G mobile wireless networks. First, we develop IRS-UAV-integrated 3D wireless channel models for mURLLC in the finite blocklength regime. Second, we formulate and solve the ϵ-effective energy-efficiency maximization problem for our proposed schemes by applying the iterative algorithms. Third, we develop the deep reinforcement learning (DRL)-based algorithms to solve the joint IRS-deployment and UAV-trajectory optimization problem. Finally, we conduct numerical analyses that validate and evaluate our developed schemes for mURLLC over IRS-UAV-relay based 6G mobile wireless networks.

Introduction

While 5G mobile wireless networks are being widely deployed around the world, researchers have begun to conceptualize 6G mobile wireless networks [1] to support unprecedented scenarios with extremely diverse and challenging QoS requirements. The delay-bounded QoS theory [2]–[4] has been proposed and developed to characterize queueing behaviors in supporting the explosively growing demands of time-sensitive and spectrum-consuming wireless multimedia applications over the emerging 6G networks, which are defined and featured, for example, in [1]. Because of the highly time-varying fading channels and complex, heterogeneous, and dynamic 6G mobile wireless network architectures, it is usually infeasible to guarantee deterministic QoS requirements for emerging multimedia traffic dominating 6G mobile wireless networks. Alternatively, *statistical QoS provisioning theory* [2]–[4], in terms of effective capacity, has been proposed as a powerful technique to characterize and implement delay-bounded QoS guarantee for wireless real-time traffic. However, with the rapid 6G developments, the exponentially increasing volumes of bandwidth-intensive and delay-sensitive multimedia traffics impose even more stringent and diverse QoS requirements, including bounded end-to-end delay (<1 ms), super-reliability (>99.99999%), and extra-high energy efficiency, etc.

Toward this end, the mURLLC [1], as one of the new and dominating 6G standard traffic services, has been proposed to quantitatively design and evaluate various 6G QoS performances. Researchers have proposed *small-packet communications* techniques, such as FBC [5], [6], in supporting various massive access techniques for reducing access latency and decoding complexity while guaranteeing the stringent QoS requirements of mURLLC [1]. The authors of [5] have shown that the codeword blocklength can be as short as 100 channel symbols for reliable communications. The authors of [6] have studied the different properties of channel codes that approach the fundamental coding rate limits using FBC.

On the other hand, QoS performance will be affected by uncontrollable interactions between transmitted radio waves and surrounding objects in dynamic wireless propagation environments. With the recent developments of *IRS* [7], [8], which consist of an array of *passive* reflecting elements to respectively impose the different phase shifts on reflected waves, network operators are able to control the scattering, reflection, and refraction characteristics of radio waves. IRSs can improve not only various QoS requirements but also radio connectivity by reducing power consumption and mitigating the stochastic nature of electromagnetic wave transmission. The authors of [9] have provided a comprehensive literature review on recent advances, applications, and design aspects of IRS. The authors of [10] have discussed the integration of IRS future Smart Cities and highlighted the potential advantages of IRS deployments.

Furthermore, since 6G wireless network frameworks are expected to provide various services combining space, aerial, and terrestrial networks for universal coverage, it is very challenging to characterize system models and guarante stringent QoS requirements in such complicated and dynamic network environments while supporting mURLLC. In the context of the small data packet transmission regime, most previous works have mainly focused on communications between ground devices and ground base stations (GBSs). However, it is not always possible to support a *massive* number of mobile devices while guaranteeing stringent mURLLC requirements through non-LOS wireless links on the ground. Inspired by the advantages of deployment capability and high mobility, *UAV* has been proposed to potentially support various *massive access techniques* by significantly enhancing LOS coverage while guaranteeing various QoS requirements.

In addition, integrating IRS with UAV provides a promising avenue to implement over-the-air intelligent reflection and enlarge wireless coverage. Motivated by the appealing advantages of *passive* IRS, IRS-UAV-integrated systems can significantly reduce energy consumption by UAVs and prolong their operational time. Compared with terrestrial IRSs, IRS-UAV-integrated systems are more likely to establish strong LOS links with ground devices

due to the relatively high altitudes and flexible 3D mobility of UAVs. Moreover, by adjusting the phase shifts of the IRS's reflecting elements, IRS-UAV-integrated systems are able to achieve panoramic full-range reflection, which significantly increases the number of supported mobile users.

There are many new challenges, including the endurance, stability, and controllability of IRS-UAV as compared with terrestrial IRS systems. In particular, it is difficult to characterize IRS-UAV-integrated 3D wireless channel models with low complexity. In addition, it is crucial to design and optimize 3D trajectories of IRS-UAV with user association to maximize system performance. Due to the complex, high-dimensional, and time-varying system state/action spaces and evolving environments, it is challenging to characterize the optimization problems for statistical delay and error-rate bounded QoS, especially when taking into account the *massive access* scenarios to support mURLLC using FBC. To address these issues, we can turn to *DRL*, which is a powerful tool for solving large-scale networking optimization problems without deriving explicit optimal solutions based on complex mathematical models. However, how to apply DRL techniques for the joint optimization of IRS deployment and UAV trajectory while statistically upper-bounding both delay and error-rate in the finite blocklength regime is an open problem over 6G wireless networks.

To effectively overcome the previously mentioned challenges, in this article, we propose FBC-based joint IRS-deployment and UAV-trajectory optimization schemes to support statistical delay and error-rate bounded QoS for mURLLC over 6G mobile wireless networks. In particular, we develop system architecture models, including a controllable IRS-UAV-integrated 3D wireless channel model and FBC-based channel coding rate model. Furthermore, we formulate and solve the FBC-based ϵ-effective energy-efficiency maximization problem by using an iterative algorithm for statistical delay and error-rate bounded QoS over IRS-UAV-integrated 6G wireless networks. We also apply the *double deep Q-network* (DDQN) algorithm to solve the proposed joint optimization problem for mURLLC. Finally, we conduct a set of simulations that validate and evaluate our developed schemes.

The Systems Architecture Models

Figure 1 shows the system architecture model for our proposed IRS-UAV-integrated 6G mobile wireless networks in supporting statistical delay and error-rate bounded QoS provisioning, which consists of one GBS, K mobile users, and one UAV, which is equipped with a large array of IRS elements to assist the communications between mobile users and the GBS. The UAV is equipped with L reflecting elements. Assume that there exists no LOS link between mobile users and the GBS,

We propose FBC-based joint IRS-deployment and UAV-trajectory optimization schemes to support statistical delay and error-rate bounded QoS for mURLLC over 6G mobile wireless networks.

and thus, the UAV operates as a passive aerial relay between mobile users and the GBS. We assume that the finite time T is equally divided into N time slots, denoted by $\mu = 1, 2, \ldots, N$, and the equal duration of each time slot is denoted by ρ. Assume that the UAV can choose to connect with K mobile users along its flight trajectory. Define the binary variable $b_k(\mu)$ for the *user association* between mobile user k and the UAV at time slot μ. When $b_k(\mu)$ is set to be 1, the UAV chooses to connect with the kth mobile user at the μth time slot. Otherwise, $b_k(\mu)$ is set to be zero.

Without loss of generality, we assume that the GBS is located at the coordinate origin. Denote by $\boldsymbol{q}_k$ the 3D-coordinate for the position of mobile users k $(k = 1, \ldots, K)$. The 3D-coordinate for the UAV's position is denoted by $\boldsymbol{q}(\mu)$ at the μth time slot with $\mu \in \{1, \ldots, \mathrm{N}\}$. We denote the UAV's flight altitude by $z(\mu)$, which must satisfy a minimum flight altitude $H_{\min}$ and maximum flight altitude $H_{\max}$, respectively. Denote by N and K the index sets for all N time slots and K mobile users, respectively.

The 3D Wireless Channel Model for the Controllable Integrated IRS-UAV

Denote $d_{k,\mathrm{U}}$ and $d_{k,\mathrm{G}}$ as the distances for the mobile user k-UAV link and UAV-GBS link, respectively. Let $\alpha_{k,\mathrm{U}} \in (0, \pi/2)$ and $\alpha_{k,\mathrm{G}} \in (0, \pi/2)$ be the elevation angles between mobile user k and the IRS-UAV node and the IRS-UAV node and the GBS, respectively, which can be determined by the UAV's flight altitude $z(\mu)$ and the transmission distances. Assume that the small-scale fading coefficient $h_{k,\mathrm{U}}^{(i)}$ between mobile user k and the ith reflecting element at the UAV follows a Rayleigh distribution. Note that due to the relative motion between the UAV and the mobile users as well as the change in the UAV's flight altitude, the UAV can only establish a proper passive reflecting link to provide a LOS path with a certain probability, which depends on environmental parameters and elevation angles.

Then, the IRS reflects the finite-blocklength signal down to the GBS through the link between the IRS-UAV and the GBS with a power reflection efficiency factor, denoted by ν, with $0 < \nu < 1$, which usually depends on the characteristics of the incident wave, the material that the metasurface is made of, and the angles of incidence and reflection. Assume that the small-scale fading coefficient $h_{k,\mathrm{G}}^{(i)}$ between the ith reflecting element at

the IRS-UAV and the GBS also follows a Rayleigh distribution. Then, the end-to-end signal-to-noise ratio (SNR) $\gamma_k^{(\mu)}$ from mobile user k to the GBS through the IRS-UAV node is a function of both the small-scale fading coefficients and transmission distances.

The Channel Coding Rate Using FBC

To support stringent QoS constraints for mURLLC services, Shannon's asymptotic capacity formalism is not applicable in the finite blocklength regime. As a result, we propose to develop an alternative analysis by applying the FBC technique to provide *statistical* delay and error-rate bounded QoS guarantees. We define a message set $\mathcal{M} = \{1, \ldots, M\}$, and a message is uniformly distributed on $\mathcal{M}$, where M is the number of codewords and $\epsilon_k^{(\mu)}$ is the decoding error probability. Correspondingly, we define an $(n, M, \epsilon_k^{(\mu)})$-code as follows:

- an encoder $\Upsilon : \{1, \ldots, M\} \mapsto \mathcal{A}^n$ that maps the message $m \in \{1, \ldots, M\}$ into a codeword with length n, where $\mathcal{A}^n$ is the codebook that represents the set of all the possible codewords mapped by the encoding function Υ
- a decoder $\mathcal{D} : \mathcal{B}^n \mapsto \{1, \ldots, M\}$ that decodes the received message into $\hat{m}$, where $\mathcal{B}^n$ is the set of received codewords of length n and $\hat{m}$ denotes the estimated signal received at the receiver; the decoder $\mathcal{D}$ needs to satisfy the maximum error probability constraint $\epsilon_k^{(\mu)}$.

Given the constrained decoding error probability, the definition expression for the *maximum achievable coding rate* $R^*(\gamma_k^{(\mu)})$ in bits per channel use with coding blocklength n from mobile user k to the GBS through the IRS-UAV node at time slot μ converges to the *outage capacity*, denoted by $C_\epsilon(\gamma_k^{(\mu)})$, as the codeword blocklength n tends to infinity.

Assume that $h_{k,\mathrm{U}}^{(i)}$ and $h_{k,\mathrm{G}}^{(i)}$ are statistically independent and identically distributed. Since $h_{k,\mathrm{U}}^{(i)}$ and $h_{k,\mathrm{G}}^{(i)}$ follow Rayleigh distribution with scale parameter σ_h, $|h_{k,\mathrm{U}}^{(i)}|^2$ and $|h_{k,\mathrm{G}}^{(i)}|^2$ follow the exponential distribution with scale parameter $1/(2(\sigma_h)^2)$. As a result, we can obtain an *upper bound* on the end-to-end SNR by using the complex Cauchy-Schwarz-Bunyakovsky inequality. Furthermore, as the number, denoted by L, of reflecting elements becomes large, we can show that the mean and variance of the product of independent Rayleigh random variables $|h_{k,\mathrm{U}}^{(i)}||h_{k,\mathrm{G}}^{(i)}|$ are $\sigma_h\pi/2$ and $4(\sigma_h)^2(1-\pi^2/16)$, respectively. Then, we can derive the *outage capacity function* via the following steps.

First, we define the *outage probability function* as the probability that the capacity of the wireless channel experiencing an outage is unable to support reliable communications at the targeted rate $R^*(\gamma_k^{(\mu)})$, implying that the SNR threshold is lower than $(2^{R^*(\gamma_k^{(\mu)})} - 1)$. Second, the outage capacity can be calculated as the maximum coding rate $R^*(\gamma_k^{(\mu)})$ such that the outage probability is

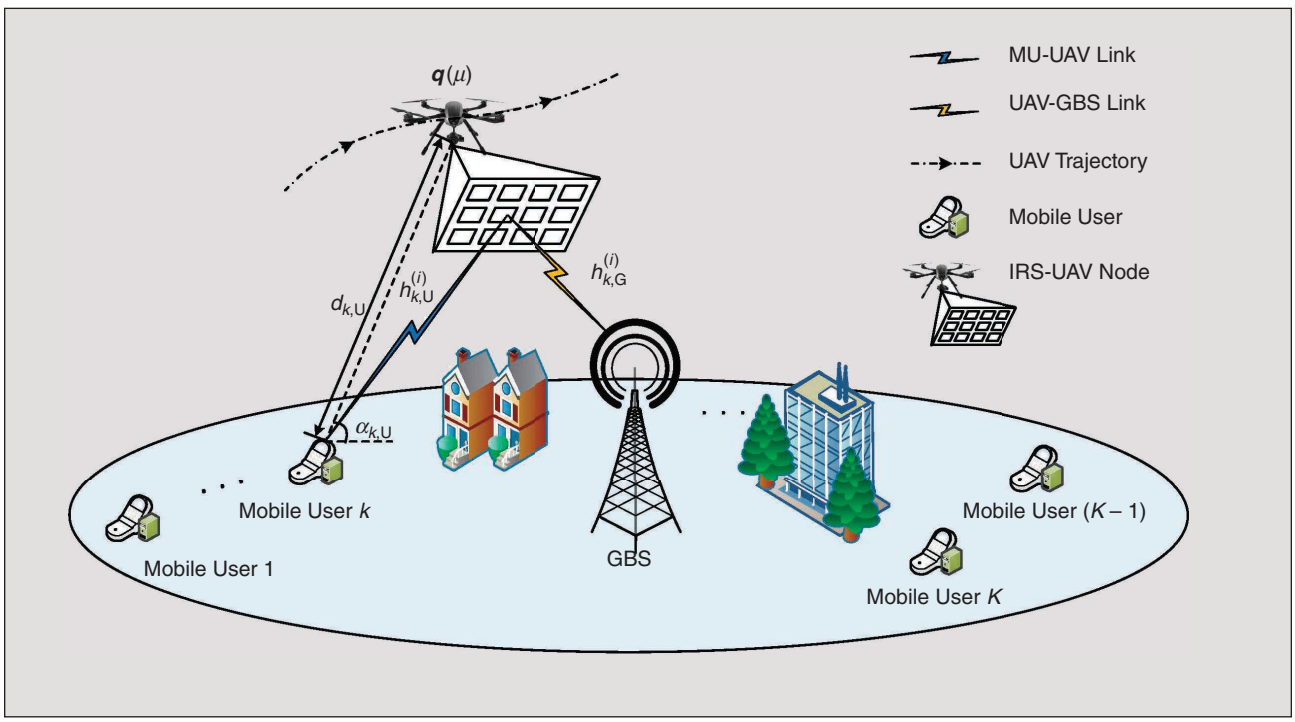


FIGURE 1 The system architecture model for our proposed IRS-UAV-integrated 6G mobile wireless networks in supporting statistical delay and error-rate bounded QoS provisioning, where $\boldsymbol{q}(\mu)$ $(\mu = 1, \ldots, N)$ is the 3D-coordinate for the position of the UAV at time slot μ; $h_{k,\mathrm{U}}^{(i)}$ and $h_{k,\mathrm{G}}^{(i)}$ $(k = 1, \ldots, K)$ $(i = 1, \ldots, L)$ are the small-scale fading coefficients between mobile user k and the ith reflecting element at the IRS-UAV node and between the ith reflecting element at the IRS-UAV node and the GBS, respectively, $d_{k,\mathrm{U}}$ and $d_{k,\mathrm{G}}$ are the link distances between mobile user k and the IRS-UAV node and the IRS-UAV node and the GBS, respectively; and $\alpha_{k,\mathrm{U}}$ is the elevation angle between mobile user k and the IRS-UAV node.

no larger than $\epsilon_k^{(\mu)}$. Third, setting the outage probability function to be equal to $\epsilon_k^{(\mu)}$, we can solve for the resulting SNR threshold for the equation by using the complementary cumulative distribution function (CCDF) of channel distribution. Fourth, since the outage probability yields the SNR threshold in terms of the maximum coding rate, this coding rate, namely, the outage capacity, can then be expressed as a function of the inverse CCDF of the channel distribution. The work in [11] also validates the previous derivations by characterizing the relationships between the outage capacity function and the CCDF of the channel distribution. Finally, defining $F_{\tilde{h}^2}(\cdot)$ as the CCDF of our channel distribution $\tilde{h}^2$, we can derive the outage capacity function as follows:

$$C_\epsilon(\gamma_k^{(\mu)}) = \log_2[1+F_{\tilde{h}^2}^{-1}(1-\epsilon_k^{(\mu)})\,\nu P_k(g_{\mathrm{U}}g_{\mathrm{G}}\sigma^2)^{-1}\times d_{k,\mathrm{U}}^{-\eta_{k,\mathrm{U}}}d_{k,\mathrm{G}}^{-\eta_{k,\mathrm{G}}}], \tag{1}$$

where $\tilde{h}$ is defined as the product of $|\Sigma_{i=1}^{L}h_{k,\mathrm{U}}^{(i)}|$ and $|\Sigma_{i=1}^{L}h_{k,\mathrm{G}}^{(i)}|$, P_k is the transmit power at mobile user k, g_{U} and g_{G} are the excessive aerial path-losses for the mobile user-UAV link and UAV-GBS link, respectively, σ^2 is the noise power, $\eta_{k,\mathrm{U}}$ and $\eta_{k,\mathrm{G}}$ are the path-loss exponents for the mobile user k-UAV link and UAV-GBS link, respectively, and $F_{\tilde{h}^2}^{-1}(\cdot)$ is the inverse of the CCDF $F_{\tilde{h}^2}(\cdot)$ with respect to $\tilde{h}^2$. Using the central limit theorem, we observe that $\tilde{h}$ approximately follows the normal distribution with mean equal to $L\sigma_h\pi/2$ and variance equal to $4L(\sigma_h)^2(1-\pi^2/16)$. Consequently, we show that $\tilde{h}^2$ follows a non-central chi-squared distribution with one degree of freedom and the non-centrality parameter $\lambda_{\tilde{h}}$. Therefore, we can derive the inverse of the CCDF $F_{\tilde{h}^2}^{-1}(\cdot)$ to obtain a closed-form expression for the outage capacity given in (1).

Joint IRS and UAV-Trajectory Optimizations Over 6G Mobile Wireless Networks Using FBC

Modeling ϵ-Effective Energy Efficiency Using FBC

We introduce the new concept of the *ϵ-effective capacity* for statistical delay and error-rate bounded QoS using FBC. Considering the nonvanishing decoding error probability $\epsilon_k^{(\mu)}$, we define the *ϵ-effective capacity*, denoted by $EC_k^{(\epsilon,\mu)}(\theta)$, from mobile user k to the GBS through the IRS-UAV node at time slot μ as the maximum constant arrival rate for a given service process subject to statistical delay and error-rate bounded QoS constraints, which is thus expressed as follows:

$$EC_k^{(\epsilon,\mu)}(\theta) \triangleq -\frac{1}{n\theta}\log\{\epsilon_k^{(\mu)} + \mathbb{E}_{\tilde{h}^2}[(1-\epsilon_k^{(\mu)}) \times \exp\{-\theta n R^*(\gamma_k^{(\mu)})\}]\}, \tag{2}$$

where $\mathbb{E}_{\tilde{h}^2}[\cdot]$ is the expectation taken with respect to $\tilde{h}^2$, θ is defined as the *QoS exponent* in units of 1/bits, which plays an important role in measuring the exponential decay rate of the delay-bounded QoS violation probabilities, and $R^*(\gamma_k^{(\mu)})$ is the maximum achievable coding rate, which can be calculated and approximated by using $C_\epsilon(\gamma_k^{(\mu)})$ given in (1).

In addition, given the limited size and power of the UAV, the number of IRS reflecting elements is limited, and the power/battery supply is also constrained for the UAV. Therefore, it is crucial to maximize the energy efficiency while optimizing the number of IRS reflecting elements for our proposed schemes. Accordingly, we define the *ϵ-effective energy efficiency*, denoted by $EE_k^{(\epsilon,\mu)}(\theta)$, as the ratio of the ϵ-effective capacity to the total power consumption, denoted by $P_{\mathrm{O},k}^{(\mu)}$, where $P_{\mathrm{O},k}^{(\mu)}$ is the total power consumption at time slot t from mobile user k to the GBS, including the following parts: 1) UAV hovering power consumption, 2) circuit power consumption, and 3) IRS hardware power consumption, denoted by P_{IRS}. Note that the IRS is acting as a passive component and does not need any transmission power. The IRS power consumption can be considered to be a function of both the power consumption of the diode in forward biased mode to operate in the ON state P_{f} and the phase resolution power consumption P_{r} [12]. Thus, the IRS power consumption is an increasing function of both the resolution and the number of IRS reflecting elements.

IRS-UAV-Integrated Joint Optimization Problem Formulation Using FBC

Our goal is to maximize the total ϵ-effective energy efficiency in the finite blocklength regime. However, since the ϵ-effective energy efficiency is in the form of the ratio of the concave and convex function with respect to the number L of IRS reflecting elements, the maximization problem can be considered as a Fractional Programming (FP) problem. Thus, we apply Dinkelbach's transform technique [13], which introduces a suitable auxiliary variable υ, to find the global optimal solution iteratively. Then, we can formulate the converted total ϵ-effective energy-efficiency maximization problem as follows:

$$\mathbf{P}_1: \underset{\substack{b_k(\mu),\,\forall\mu\in\mathsf{N},\forall k\in\mathsf{K}\\ L,\,\boldsymbol{q}(\mu),\,\forall\mu\in\mathsf{N}}}{\operatorname{argmax}} \left\{\sum_{\mu=1}^{N}\sum_{k=1}^{K}[EC_k^{(\epsilon,\mu)}(\theta) - \upsilon P_{\mathrm{O},k}^{(\mu)}]\right\}, \tag{3}$$

subject to the UAV's trajectory constraints, including flight altitude constraint and flight velocity constraint, as well as the IRS's reflecting elements constraint. Note that $\mathbf{P}_1$ is a *mixed-integer nonconvex* optimization problem because the user association indicator is a binary variable and the number of IRS reflecting elements is an *integer* variable, while the UAV's trajectory is a *continuous* variable. Thus, we cannot directly solve $\mathbf{P}_1$ by using standard convex optimization techniques. Therefore, we divide the optimization problem $\mathbf{P}_1$ into the following three subproblems and solve them in an iterative manner.

1) Given the number L of IRS reflecting elements and UAV trajectory $\boldsymbol{Q}$, $\mathbf{P_1}$ can be converted into a convex optimization problem in terms of the user association $\boldsymbol{B}$. Then, there exists a unique optimal solution, and the global optimum can be efficiently found by applying the combination of a traditional convex optimization algorithm with an exhaustive search, which can quickly and efficiently find the optimal solution when the limits are set appropriately. This can be done by applying the CVX toolbox [14].
2) Given the number L of IRS reflecting elements and user association vector $\boldsymbol{B}$, $\mathbf{P_1}$ can be converted into a convex optimization problem in terms of the UAV trajectory $\boldsymbol{Q}$ and solved by using the CVX toolbox.
3) Given the UAV trajectory $\boldsymbol{Q}$ and user association vector $\boldsymbol{B}$, $\mathbf{P_1}$ can be converted into a convex optimization problem in terms of the number of IRS reflecting elements and solved by using the CVX toolbox.

We develop an iterative algorithm as detailed in **Algorithm 1** to solve $\mathbf{P_1}$ for our proposed schemes using FBC. Particularly, we can solve $\mathbf{P_1}$ in an iterative manner until it converges to a prespecified accuracy. We define $\boldsymbol{B}^{(\ell)}$, $\boldsymbol{Q}^{(\ell)}$, $L^{(\ell)}$, and $\upsilon^{(\ell)}$ as the user association vector, 3D UAV location vector, the number of IRS reflecting elements, and the auxiliary variable, respectively, at the ℓth iteration. Define ς as the iteration tolerance threshold. In each iteration, the complexity of **Algorithm 1** is dominated by Step 1 and Step 2. The complexity of Step 1 is $O(KN)$. The complexity of Step 2 is $O((KN)^{3.5})$ by applying the successive convex approximation (SCA) technique and the CVX toolbox. Accordingly, the total complexity of **Algorithm 1** is $O(I_{\text{iter}}(KN+(KN)^{3.5}))$, where I_{iter} is the number of iterations required to meet a stopping criterion.

The DRL-Based Joint Optimization for mURLLC Over IRS-UAV-Integrated Wireless Networks Using FBC

In the previous section, we have formulated and solved the ϵ-effective energy-efficiency maximization problem in the finite blocklength regime using iterative algorithm.

ALGORITHM 1 *Iterative Algorithm for Solving* $\mathbf{P_1}$.

Input: K, N, n, and iteration tolerance threshold ς;
Initialization: $\boldsymbol{Q}^{(0)}, L^{(0)}, \boldsymbol{B}^{(0)}$, and $\upsilon^{(0)}$
Repeat
Step 1: Given $\boldsymbol{Q}^{(\ell)}$ and $L^{(\ell)}$, update $\boldsymbol{B}^{(\ell+1)}$ by using the CVX toolbox.
Step 2: Given $\boldsymbol{B}^{(\ell+1)}$ and $L^{(\ell)}$, update $\boldsymbol{Q}^{(\ell+1)}$ by using the CVX toolbox.
Step 3: Given $\boldsymbol{Q}^{(\ell+1)}$ and $\boldsymbol{B}^{(\ell+1)}$, update $L^{(\ell+1)}$ by using the CVX toolbox.
Update the auxiliary variable $\upsilon^{(\ell+1)}$.
$\ell \leftarrow (\ell+1)$.
Until The fractional increase in the reformulated objective function in $\mathbf{P_1}$ is no larger than the iteration tolerance threshold ς.

However, considering dynamic wireless propagation environments, the optimization for $\mathbf{P_1}$ tends to quickly become infeasible, especially for *massive access* scenarios. The time efficiency of iterative algorithms may be poor, and the computational complexity will increase dramatically as the problem scale becomes larger and conventional iterative algorithms may not be feasible. As a result, in this section applying the Markov Decision Process (MDP) methodology, we apply DRL, which can exert different actions under different network states with an *infinite* number of trials until it can gradually adapt to the dynamically varying environments according to feedback received in real time.

MDP Formulation for DRL

We apply an MDP, which associates an action to each network state, denoted by $\boldsymbol{s}(\mu)$, for selecting a joint optimal policy. An MDP can be described by a 5-tuple $(\mathsf{S}, \mathsf{A}, \mathsf{P}, \mathsf{R}, \beta)$, where S is the state space containing all possible states; A is the action space collecting all possible actions, denoted by $\boldsymbol{a}(\mu)$, P, represents the transition probabilities; $\Pr\{\boldsymbol{s}(\mu+1)\,|\,\boldsymbol{s}(\mu), \boldsymbol{a}(\mu)\}$, R is a reward fed back to the agent after executing an action; and $\beta \in [0,1)$ represents the discount factor, determining the importance of long-term rewards. To solve $\mathbf{P_1}$, we define the above mentioned elements in an MDP as follows:

- **State**: The state $\boldsymbol{s}(\mu) \in \mathsf{S}$ consists of two parts: the UAV's 3D locations $\boldsymbol{q}(\mu)$ and the small-scale fading coefficients $\{h_{k,\mathrm{U}}^{(i)}, \forall k \in \mathsf{K}, i \in \mathsf{L}\}$ and $\{h_{k,\mathrm{G}}^{(i)}, \forall k \in \mathsf{K}, i \in \mathsf{L}\}$ for the mobile user-UAV link and UAV-GBS link, respectively.
- **Action**: We define $\boldsymbol{a}(\mu) \in \mathsf{A}$ as the action for the UAV trajectory $\boldsymbol{Q}$ and user association for all K mobile users.
- **Reward function**: The agent can receive a reward, denoted by $R(\mu+1)$, along with the state $\boldsymbol{s}(\mu+1)$. Since our goal is to maximize the ϵ-effective energy efficiency for supporting mURLLC constraints, the reward function $R(\mu)$ at time slot μ can be defined as the ϵ-effective energy efficiency.

Optimal DDQN-Based Joint Optimization Algorithm Using FBC

We propose to develop the DDQN-based algorithm to solve the joint optimization problem through trial-and-error interactions with the wireless propagation environments. By applying the expected discounted cumulative reward function, the DRL agent updates its Q-function in an online manner to progressively find the optimal IRS and UAV trajectory strategy for each state $\boldsymbol{s}(\mu)$. The key to the DDQN algorithm is to select an action by using the primary network, and then, the target network is used to calculate the target Q-value for the action. The DDQN agent learns a state-action value function approximator, denoted by $Q(\boldsymbol{s}(\mu), \boldsymbol{a}(\mu)|\omega_{\text{DDQN}}^{(\mu)})$, where $\boldsymbol{\omega}_{\text{DDQN}}$ is the weight matrix of the DDQN, which is updated in a fully

online manner to avoid the complexities of eligibility traces. Using a DDQN training principle [15], we can apply the gradient of the loss function, denoted by $\nabla\mathcal{L}(\boldsymbol{\omega}_{\text{DDQN}}^{(\mu)})$, to train the value function approximator, where ∇ is the vector differential operator.

Let $\psi \in (0,1)$ be the updating rate for the DDQN algorithm and I be the total number of operation iterations. A greedy policy is applied to avoid overfitting. At each training iteration, it either chooses the best available action in a given state with probability $(1-\tau)$ or samples a random action with probability τ. The gradient of the loss function $\nabla\mathcal{L}(\boldsymbol{\omega}_{\text{DDQN}}^{(\mu)})$ is calculated with respect to a random minibatch, which is uniformly chosen at random from a finite replay memory, denoted by Γ. To remove the degree of correlation among the observed sequence of data and improve the stability of DDQN, we adopt the experience replay approach, where the system transition 4-tuple $(\boldsymbol{a}(\mu), \boldsymbol{s}(\mu), \boldsymbol{s}(\mu+1), R(\mu+1))$ is stored in the replay memory Γ after each iteration. At each iteration, a minibatch of the transition tuple is randomly drawn from the replay memory Γ, and batch gradient descent is employed to minimize the loss functions of the minibatch of the transition tuple. Thus, the previous experiences are exploited more efficiently as the algorithm can learn from them many times. We develop a DDQN-based algorithm, as shown in **Algorithm 2,** to solve the maximization problem $\mathbf{P_1}$ for our proposed schemes. The computational complexity of our proposed DDQN-based algorithm is $Q(N|\mathsf{S}|^2\mathsf{A})$.

Performance Evaluations

We provide a set of numerical results to validate and evaluate our proposed IRS-UAV-integrated joint optimization schemes for statistical delay and error-rate bounded QoS over 6G mobile wireless networks. Throughout our simulations, we set the number of mobile users $K = 200$, the maximum height of the UAV $H_{\max} = 800$ m, the minimum height of the UAV $H_{\min} = 30$ m, the scale parameter of the Rayleigh distribution $\sigma_h = 1/\sqrt{2}$, and the excessive aerial path losses $g_{\text{U}} = g_{\text{G}} = 0.002$.

Setting the blocklength $n = 500$ and the QoS exponent $\theta = 1\times10^{-4}$, Figure 2 depicts the average ϵ-effective capacity as a function of the number of IRS reflecting elements using FBC. We can observe from Figure 2 that the average ϵ-effective capacity is a monotonically increasing function with respect to the number of IRS reflecting elements. Figure 2 also shows that the average ϵ-effective capacity is a monotonically decreasing function in terms of the distance.

We now set the decoding error probability $\epsilon_k^{(\mu)} \in \{0.1, 0.01\}$ and the number of IRS reflecting elements $L = 300$. Figure 3 plots the average ϵ-effective capacity as a function of the QoS exponent using FBC. Figure 3 shows that the average ϵ-effective capacity is a monotonically decreasing function of the QoS exponent. We can also observe from Figure 3 that given a decoding error probability, the gap between the curves for ϵ-effective capacity with different distances becomes smaller as the QoS

***Algorithm 2** DDQN-Based Algorithm for Joint Optimization of IRS and UAV Trajectory.*

Input: K, N, n, β, I, and a threshold τ;
Initialization: The action-state value function $Q(\boldsymbol{s}, \boldsymbol{a}\,|\,\omega_{\text{DDQN}})$
for $i = 1, \ldots, I$ **do**
  Set $\ell = 1$
  Initialize the environment and receive an initial state $\boldsymbol{s}(1)$.
  for $\mu = 1, \ldots, N$ **do**
    if $rand(\cdot) < \tau$ **then**
      Select a random action from A.
    else
      Observe the current state $\boldsymbol{s}(\mu)$, and select an action $\boldsymbol{a}(\mu) = \max_{\boldsymbol{a}\in\mathsf{A}}\{Q(\boldsymbol{s}(\mu), \boldsymbol{a}\,|\,\omega_{\text{DDQN}})\}$.
    end if
    The GBS observes $\boldsymbol{s}(\mu+1)$ and calculates the immediate reward and stores the transition $(\boldsymbol{a}(\mu), \boldsymbol{s}(\mu), \boldsymbol{s}(\mu+1), \boldsymbol{R}(\mu+1))$ in replay memory Γ.
    Randomly sample a minibatch of transitions from replay memory Γ.
    Calculate the loss function.
    Perform a gradient descent for each primary network and update the target networks using $\bar{\omega}_{\text{DDQN}}^{(\mu)} \leftarrow \psi\omega_{\text{DDQN}}^{(\mu)} + (1-\psi)\bar{\omega}_{\text{DDQN}}^{(\mu)}$.
  end for
end for

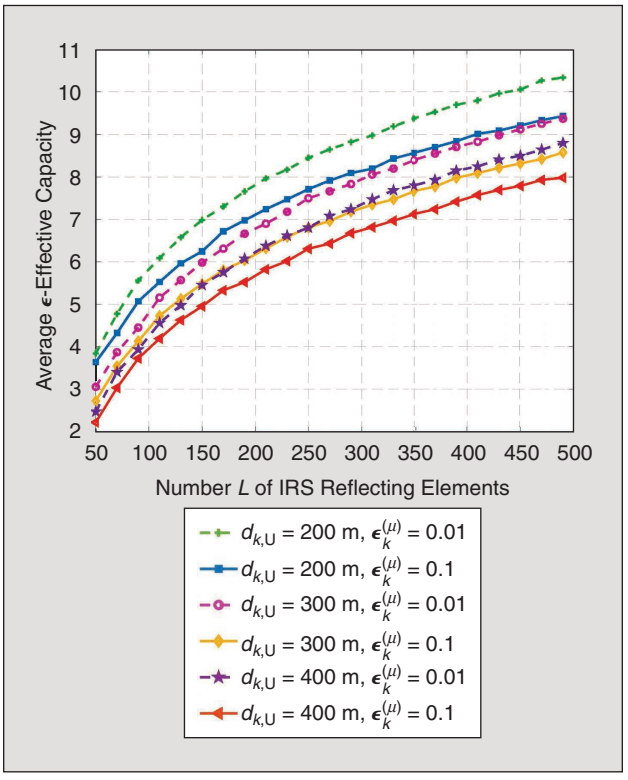


Figure 2 The average ϵ-effective capacity versus the number L of IRS reflecting elements for our developed schemes using FBC with the decoding error probability $\epsilon_k^{(\mu)} \in \{0.1, 0.01\}$.

exponent increases. This implies that with a large QoS exponent, the delay-bounded QoS constraints become very stringent, which leads to a very small, achievable ϵ-effective capacity. Therefore, increasing the distance does not significantly affect the value of very small ϵ-effective capacity.

Setting the blocklength $n = 500$ and the phase resolution power consumption $P_r = 0.01$ W, Figure 4 depicts the average ϵ-effective energy efficiency as a function of the UAV flight altitude. We can observe from Figure 4 that the average ϵ-effective energy efficiency first increases and then decreases as the UAV flight altitude increases, which reveals that there exists an optimal UAV flight altitude that maximizes the ϵ-effective energy efficiency for our proposed schemes. Figure 4 also shows that the average ϵ-effective energy efficiency decreases as the QoS exponent increases, implying that a smaller θ and a larger θ set an upper bound and lower bound on the ϵ-effective energy efficiency, respectively.

Setting the QoS exponent $\theta = 1 \times 10^{-3}$ and decoding error probability $\epsilon_k^{(\mu)} = 0.01$, Figure 5 plots the average ϵ-effective energy efficiency as a function of both the number of IRS reflecting elements and the phase resolution power consumption using FBC. We can observe from Figure 5 that the ϵ-effective energy efficiency increases as the phase resolution power consumption decreases. Figure 5 also shows that the ϵ-effective energy efficiency first increases and then decreases as the number of IRS reflecting elements increases, which reveals the importance of optimizing the number of IRS reflecting elements to maximize the ϵ-effective energy efficiency.

Now we set the blocklength $n = 500$, the QoS exponent $\theta = 1 \times 10^{-3}$, and phase resolution power consumption $P_r = 0.01$ W. Using **Algorithm 2**, Figure 6 depicts the

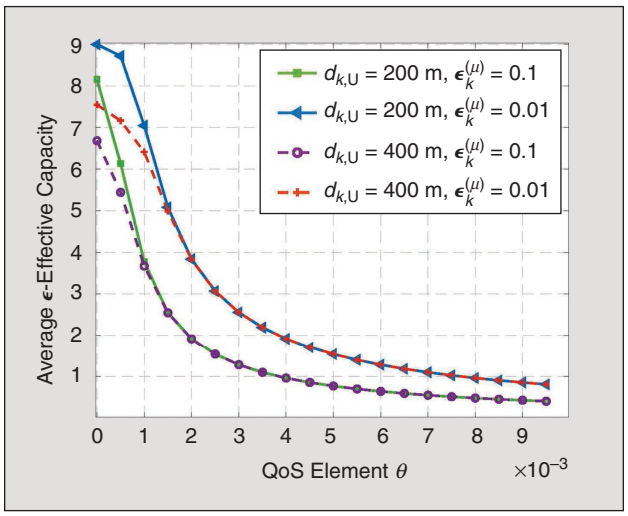


FIGURE 3 The average ϵ-effective capacity versus the QoS exponent θ for our developed schemes using FBC.

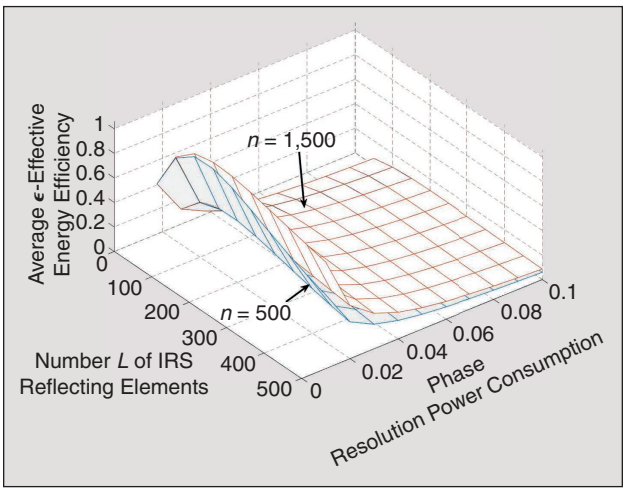


FIGURE 5 The average ϵ-effective energy efficiency versus the number of IRS reflecting elements and phase resolution power consumption P_r for our developed schemes using FBC.

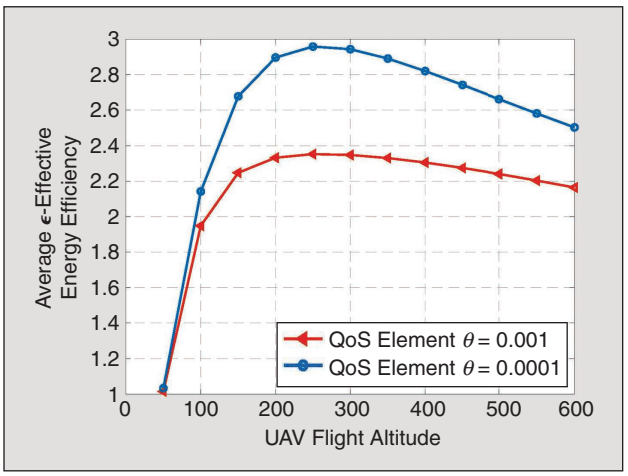


FIGURE 4 The average ϵ-effective energy efficiency versus the UAV flight altitude for our developed schemes using FBC with the QoS exponent $\theta \in \{1 \times 10^{-3}, 1 \times 10^{-2}\}$.

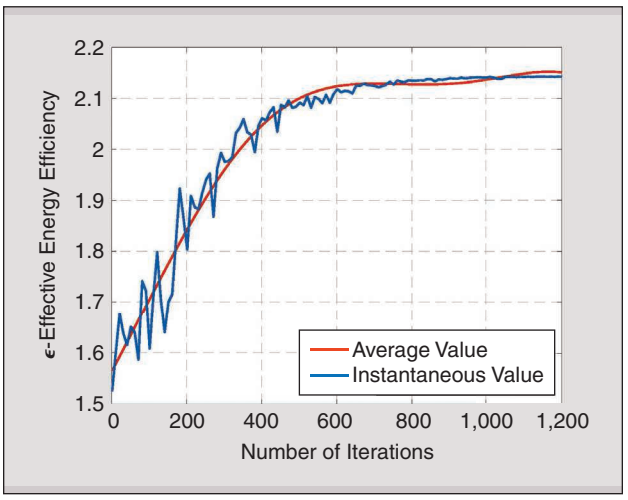


FIGURE 6 The average ϵ-effective energy efficiency versus the number of iterations for our developed DDQN-based algorithm using FBC.

ϵ-effective energy efficiency against the number of training iterations for both instantaneous and average values to verify the convergence of our proposed DDQN-based joint optimization algorithm. We also observe from Figure 6 that after around 500 iterations, the system starts to converge.

Conclusions

We have developed the joint optimization of IRS-deployment and UAV-trajectory for statistical delay and error-rate bounded QoS provisioning schemes over 6G mobile wireless networks using FBC. In particular, we have established an IRS-UAV-integrated 3D wireless network communication model and FBC-based channel coding model. We have formulated and solved the FBC-based ϵ-effective energy-efficiency maximization problem by using the iterative algorithm for statistical delay and error-rate bounded QoS over IRS-UAV-integrated wireless networks. We also have proposed the DDQN-based algorithm to solve the joint optimization problem for mURLLC. Finally, we have conducted a set of numerical analyses to validate and evaluate our developed schemes over IRS-UAV-integrated 6G mobile wireless networks.

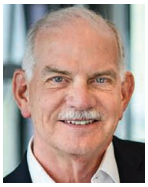
Acknowledgments

The work of Xi Zhang and Jingqing Wang was supported in part by the U.S. National Science Foundation under Grants CCF-2008975, ECCS-1408601, and CNS-1205726 and the U.S. Air Force under Grant FA9453-15-C-0423. The work of H. Vincent Poor was supported in part by the U.S. National Science Foundation under Grant CCF-1908308. Xi Zhang is the corresponding author of this article.

Author Information

Xi Zhang (xizhang@ece.tamu.edu) is a full professor with the Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, 77843, USA. He received his Ph.D. degree in electrical engineering and computer science from The University of Michigan, Ann Arbor, USA. He has published more than 400 research papers. He received the U.S. NSF CAREER Award in 2004 and six IEEE Best Paper awards. He is an IEEE Distinguished Lecturer for both IEEE ComSoc and IEEE VTS. He is an editor for numerous IEEE transactions and journals. He is a Fellow of IEEE.

Jingqing Wang (wang12078@tamu.edu) is with the Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas, 77843, USA. She received her B.S. degree from Northwestern Polytechnical University, Xi'an, China, in electronics and information engineering and is currently pursuing her Ph.D. degree under the supervision of Professor Xi Zhang in Networking and Information Systems Laboratory, Department of Electrical and Computer Engineering, Texas A&M University, College Station, Texas.



H. Vincent Poor (poor@princeton.edu) is the Michael Henry Strater University Professor at Princeton University, Princeton, New Jersey, 08544, USA, where his interests lie in wireless networks, energy systems, and related fields. Among his publications is the forthcoming book *Machine Learning and Wireless Networks* (Cambridge University Press). A member of the U.S. National Academies of Engineering and Sciences, he received the IEEE Vehicular Technology Society (VTS) Hall of Fame Award in 2021 and is currently serving as a VTS Distinguished Lecturer. He is a Life Fellow of IEEE.

References

[1] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *IEEE Netw.*, vol. 34, no. 3, pp. 134–142, May/Jun. 2020, doi: 10.1109/MNET.001.1900287.

[2] X. Zhang, J. Tang, H.-H. Chen, S. Ci, and M. Guizani, "Cross-layer-based modeling for quality of service guarantees in mobile wireless networks," *IEEE Commun. Mag.*, vol. 44, no. 1, pp. 100–106, 2006, doi: 10.1109/MCOM.2006.1580939.

[3] J. Tang and X. Zhang, "Quality-of-service driven power and rate adaptation over wireless links," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 3058–3068, 2007, doi: 10.1109/TWC.2007.051075.

[4] H. Su and X. Zhang, "Cross-layer based opportunistic MAC protocols for QoS provisionings over cognitive radio wireless networks," *IEEE J. Sel. Areas Commun. (J-SAC)*, vol. 26, no. 1, pp. 118–129, Jan. 2008.

[5] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010, doi: 10.1109/TIT.2010.2043769.

[6] Y. Polyanskiy and S. Verdú, "Empirical distribution of good channel codes with nonvanishing error probability," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 5–21, 2013, doi: 10.1109/TIT.2013.2284506.

[7] Q. Wu and R. Zhang, "Towards smart and reconfigurable environment: Intelligent reflecting surface aided wireless network," *IEEE Commun. Mag.*, vol. 58, no. 1, pp. 106–112, 2020, doi: 10.1109/MCOM.001.1900107.

[8] C. Pan *et al.*, "Intelligent reflecting surface aided MIMO broadcasting for simultaneous wireless information and power transfer," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1719–1734, Aug. 2020, doi: 10.1109/JSAC.2020.3000802.

[9] S. Gong *et al.*, "Toward smart wireless communications via intelligent reflecting surfaces: A contemporary survey," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2283–2314, 2020, doi: 10.1109/COMST.2020.3004197.

[10] S. Kisseleff, W. A. Martins, H. Al-Hraishawi, S. Chatzinotas, and B. Ottersten, "Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities," *IEEE Open J. Commun. Soc.*, vol. 1, pp. 1781–1797, Nov. 2020, doi: 10.1109/OJCOMS.2020.3036839.

[11] A. S. Avestimehr and D. N. C. Tse, "Outage capacity of the fading relay channel in the low-SNR regime," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1401–1415, 2007, doi: 10.1109/TIT.2007.892773.

[12] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Trans. Wireless Commun.*, vol. 18, no. 8, pp. 4157–4170, Aug. 2019, doi: 10.1109/TWC.2019.2922609.

[13] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, 1967, doi: 10.1287/mnsc.13.7.492.

[14] M. Grant and S. Boyd, "CVX: MATLAB software for disciplined convex programming," CVX Research, Version 2.2, Mar. 2014. [Online]. Available: http://cvxr.com/cvx

[15] T. P. Lillicrap *et al.*, "Continuous control with deep reinforcement learning," 2015, *arXiv:1509.02971*.

VT