Risk-Based Optimization of Virtual Reality over Terahertz Reconfigurable Intelligent Surfaces

Christina Chaccour*, Mehdi Naderi Soorki†, Walid Saad *, Mehdi Bennis‡, and Petar Popovski §
*Wireless@ VT, Bradly Department of Electrical and Computer Engineering, Virginia Tech, Blacksburg, VA USA,
†Faculty of Engineering, Shahid Chamran University of Ahvaz, Ahvaz, Iran,
‡Centre for Wireless Communications, University of Oulu, Finland,
§ Department of Electronic Systems, Aalborg University, Denmark.
Emails:{christinac, mehdin, walids}@vt.edu, mehdi.bennis@oulu.fi, petarp@es.aau.dk

***Abstract*—In this paper, the problem of associating reconfigurable intelligent surfaces (RISs) to virtual reality (VR) users is studied for a wireless VR network. In particular, this problem is considered within a cellular network that employs terahertz (THz) operated RISs acting as base stations. To provide a seamless VR experience, high data rates and reliable low latency need to be continuously guaranteed. To address these challenges, a novel risk-based framework based on the entropic value-at-risk is proposed for rate optimization and reliability performance. Furthermore, a Lyapunov optimization technique is used to reformulate the problem as a linear weighted function, while ensuring that higher order statistics of the queue length are maintained under a threshold. To address this problem, given the stochastic nature of the channel, a policy-based reinforcement learning (RL) algorithm is proposed. Since the state space is extremely large, the policy is learned through a deep-RL algorithm. In particular, a recurrent neural network (RNN) RL framework is proposed to capture the dynamic channel behavior and improve the speed of conventional RL policy-search algorithms. Simulation results demonstrate that the maximal queue length resulting from the proposed approach is only within 1% of the optimal solution. The results show a high accuracy and fast convergence for the RNN with a validation accuracy of 91.92%.**

Index Terms— Virtual Reality, Terahertz, Reliability

I. Introduction

Virtual reality (VR) applications will revolutionize the way in which humans interact by allowing them to be immersed, in real time, in a range set of virtual environments [1]. Nevertheless, unleashing the potential of VR requires their integration into wireless networks in order to provide a seamless and immersive VR experience [2]. However, deploying wireless VR services faces many technical challenges, the most fundamental of which is providing high rate wireless links with high reliability. On the one hand, VR communication requires high data rates to guarantee a seamless visual experience while delivering 360° VR content. On the other hand, providing reliable haptic VR communications will also require maintaining a very low end-to-end (E2E) delay in face of extreme and uncertain network conditions.

Guaranteeing this dual performance requirement constitutes a major departure from classical ultra reliable low latency communications (URLLC) services limited to low-rate sensors [3], or traditional enhanced mobile broadband (eMBB) services limited to high capacity delivered to dense networks [4]. In order to overcome the rate challenge of VR communications, one can explore the high bandwidth available at the terahertz (THz) frequency bands [5]. However, the reliability of the THz channel can be impeded by its susceptibility to blockage, molecular absorption, and communication range. This, in turn, can violate the reliability requirements of VR systems. In order to alleviate these reliability concerns, one can deploy reconfigurable intelligent surfaces (RISs) [6] acting as a base station (BS) that can provide a nearly continuous line-of-sight (LoS) connectivity to VR users. The RIS concept can be viewed as a scaled-up version of conventional multiple-input multiple-output (MIMO) systems beyond their traditional large array concept, however, an RIS exhibits several key differences from massive MIMO systems [6]. Most fundamentally, RISs will be densely located in both indoor and outdoor spaces, making it possible to perform near-field communications through a line-of-sight (LoS) path. Hence, coupling RISs with THz communications can potentially provide connectivity that exhibits both high data rates and high reliability (in terms of guaranteeing LoS communication). Moreover, VR users will always be at a proximity of physical structures with high rate wireless capabilities. Thus, it is imperative to understand whether THz-operated RISs can indeed provide an immersive VR experience by delivering continuously reliable connectivity with low E2E delay and high data rate.

A number of recent works attempted to address the challenges of VR communications [1], [7]–[9]. In [7], the authors study the spectrum resource allocation problem with a brain-aware quality-of-service (QoS) constraint. The work in [1] proposes a VR model that captures the tracking and delay components of VR QoS. Meanwhile, the work in [8] proposes a novel framework that uses cellular-connected drone aerial vehicles to collect VR content for reliable wireless transmission. In [9] the authors study the issue of concurrent support of visual and haptic perceptions over wireless cellular networks. However, the works in [1] and [7]–[9] do not account for realistic delays and their statistics, and their solutions cannot satisfy high rates and low latency simultaneously. In contrast, to provide reliable VR, it is of interest to explore the possibility of deploying RISs. If properly operated, serving VR users through existing walls and structures with wireless capabilities will

This research was supported by the U.S. National Science Foundation under Grant CNS-1836802, and in part, by the Academy of Finland Project CARMA, by the Academy of Finland Project MISSION, by the Academy of Finland Project SMARTER, as well as by the INFOTECH Project NOOR.

unleash the potential of reliable VR. Specifically, equipping RISs with THz will guarantee the overall seamless experience. We also note that despite the surge of recent works on THz communications (e.g., see [10] and [5], and references therein) and RIS design and optimization (e.g., see [11], [12], and references therein), these works focus on the physical layer and do not address VR or networking challenges of THz communications.

The main contribution of this paper is a novel rate and reliability optimization framework for VR systems leveraging THz-operated RISs. We consider the downlink of a cellular network in which THz-operated RISs serve VR users. In this network, due to the mobility of users and the stochastic nature of the channel, the RISs must be dynamically and intelligently scheduled to VR users. Also, to guarantee reliability and capture a full knowledge of delay statistics, we propose a novel approach that exploits the economic concept of entropic value-at-risk (EVaR) which *coherently* measures the risk associated to a random event [13]. Hence, the EVaR is employed so as to capture higher order statistics of the delay and, thus, allowing us to define a concrete a measure of the risk associated to delay unreliability. We then formulate a high reliability and sum rate maximization scheduling problem by combining both Lyapunov optimization and deep neural networks (DNNs). Using the Lyapunov optimization technique, the problem is transformed into a linear weighted function, which ensures that the maximum queue length and the maximum queue length variance among VR users remains bounded. To solve the proposed Lyapunov optimization problem, we propose a reinforcement learning (RL) algorithm based on recurrent neural networks (RNNs) that can find the user associations to RISs while capturing the dynamic temporal behavior of the users in the channel. Simulation results show that the gap between the proposed approach and the optimal solution is minimal.

The rest of this paper is organized as follows. The system model is presented in Section II. The risk aware association for VR users is proposed in Section III. The RL approach is presented Section IV. In Section V, we provide simulation results. Finally, conclusions are drawn in Section VI.

II. System Model

Consider the downlink[1] of an RIS-based wireless network in a confined indoor area, servicing a set $\mathcal{U}$ of U mobile wireless VR users via a set $\mathcal{B}$ of B RISs acting as THz operated BSs as depicted in Fig. 1. The VR users are mobile and may change their locations and orientations at any point in time. We consider discrete time slots indexed by t with fixed duration τ. Each RIS is a BS, that is provided with a feeder (antenna) with a corresponding transmit power denoted by p. Hence, the transmitted data is encoded onto the phases of the signals reflected from different reconfigurable meta-surfaces that compose the RIS [6]. Henceforth, if the RIS consists

[1]The uplink of VR requests is assumed to follow an arbitrary URLLC scheme and is outside of the scope of this paper

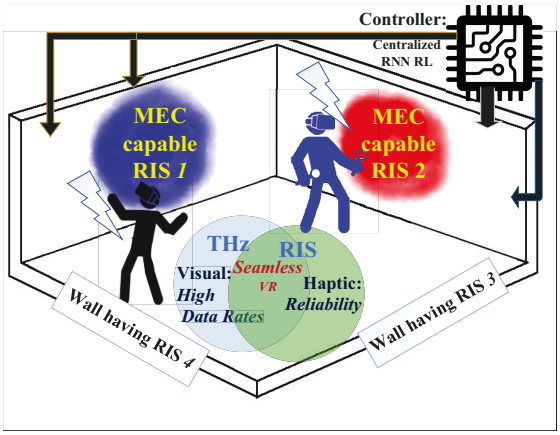

Fig. 1: Illustrative example of our system model.

of N meta-surfaces whose reflection phase can be optimized independently, then an N-stream virtual MIMO system can be realized by using a single radio frequency (RF) active chain [6]. We assume that the RF source is close enough to the RIS surface so that the transmission between each pair of RF source and RIS is not affected by fading. Then, the electromagnetic response of the N meta-surface elements can be programmed by using a centralized controller, which generates input signals that tune varactors and change the phase of the reflected signal [11]. Let $\Phi_{bu,t} = [\phi_{bun,t}]_{N\times 1}$ be the phase shift vector of RIS b , with respect to the user equipment (UE) u, at time slot t, where $\phi_{bun,t} \in \Phi$, n is the index corresponding to the meta-surface of each RIS, and $\Phi = \{-\pi + \frac{2z\pi}{Z-1} | z = 0, 1, ..., Z-1\}$. Z is the number of possible phase shifts per meta-surface element.

A. Channel Model

Due to the mobility of the VR users, the THz link between a VR user and its respective RISs may be blocked by self-blockage, i.e., the event of blocking the signal received by UE u's own body, or by dynamic blockages, i.e., blocking the signal by other VR users' bodies respectively. Let $s_{bu,t}$ be a random binary variable where $s_{bun,t} = 1$ if there is a LoS link between RIS b and VR UE u at time slot t, and $s_{bun,t} = 0$, otherwise. As a byproduct of directional beamforming and propagation differences, the network considered is noise limited. Thus, the random channel gain between RIS b and VR UE u at time slot t is given by [14]:

$$h_{bu,t} = \begin{cases} \left(\frac{\lambda}{4\pi d_{bu,t}}\right)^2 \left(e^{-k(f)d_{bu,t}}\right)^2, & \text{with } \Pr(s_{bun,t} = 1), \\ 0, & \text{with } \Pr(s_{bun,t} = 0). \end{cases}$$

$d_{bu,t}$ is the distance between RIS b and the VR UE u at time slot t, $k(f)$ is the overall molecular absorption coefficients of the medium at THz band, and f is the operating frequency. Let $\psi_{bun,t}$ be the phase shift of the channel between VR UE u and the meta-surface n of RIS b at t. Then, for a given reflection phase shift vector, $\Phi_{bu,t}$, the transmission rate from RIS b to VR UE u will be (under an approximate average signal-to-noise ratio (SNR) value across the THz band):

$$\tilde{r}_{bu,t} = W \log_2 \left(1 + \frac{p h_{bu,t} |\sum_{n=1}^{N} e^{(\phi_{bun,t} - \psi_{bun,t})j}|^2 s_{bun,t}}{N(d_{bu,t}, p, f)} \right), \quad (1)$$

where $N(d_{u,t}, p, f) = N_0 + \sum_{b=1}^{B} p A_0 d_{bu,t}^{-2} (1 - e^{-K(f) d_{bu,t}})$, $N_0 = \frac{W\lambda^2}{4\pi} k_B T_0$, k_B is the Boltzmann constant, T_0 is the temperature in Kelvin, $A_0 = \frac{c^2}{16\pi^2 f^2}$, and c is the speed of light [14], [15]. Note that the optimal choice for $\phi_{bun,t}$ for every RIS association is equal to $\psi_{bun,t}$, thus maximizing the rate $\tilde{r}_{bu,t}$, as shown in [6]. This selection will be made by the controller after learning the RIS association and optimizing it, as shown in subsequent sections.

B. Queuing Model

Each RIS is equipped with mobile edge computing (MEC) capabilities, and, thus, we model the queuing and transmission of each VR content as an M/G/1 queue at the MEC server of the RIS. We define a decision binary variable $x_{bu,t}$ that is equal to 1 if RIS b is scheduled to serve the VR content queue of user u at time slot t, otherwise $x_{bu,t} = 0$. Note that, multiple users can be associated to one RIS, however, each user is connected to a single RIS. Let $Q_u(t)$ be the queue length corresponding to UE u's requested VR image at the beginning of slot t, then the queue dynamics are given by:

$$Q_{u,(t+1)} = \max\{Q_{u,t} - \tilde{R}_{bu,t}, 0\} + A_{u,t}, \text{if } x_{bu,t} = 1, \quad (2)$$

where $A_{u,t}$ is the number of VR images queued for transmission at time slot t. The arrival of VR content follows a Poisson arrival process with mean rate λ_u. $\tilde{R}_{bu,t}$ is the rate of VR image transmission over THz link between RIS b and VR UE u at time slot t. $\tilde{R}_{bu,t} = \frac{\tilde{r}_{bu,t}\tau}{M}$ where M is the size of the VR image. Given that the availability of the THz LoS link is a random variable, $\tilde{R}_{bu,t}$ is a stochastic random variable with respect to time.

III. Risk-Aware RIS-VR user Association

A. Problem Formulation

Our goal is to characterize the RIS-UE association policy which determines the system parameters over a finite horizon of length T. The objective of this optimal policy is to maximize the sum-rate while maintaining reliable transmission. Formally, we define a *policy* $\Pi_t = \{x_{bu,t} | \forall b \in \mathcal{B}, \forall u \in \mathcal{U}\}$ for the controller that associates each RIS to its respective VR users. The control policy at a given slot t depends on unknown environmental changes, which is a consequence of the stochastic nature of the channel and the sudden changes that might block the LoS signal between RISs and mobile VR UE. Thus, $\Pr(s_{bun,t} = j | s_{bun,(t-1)} = i), \forall b \in \mathcal{B}, \forall u \in \mathcal{U}$, over the LoS THz links. Furthermore, the reliability metric is satisfied as long as the cumulative distribution function (CDF) of the E2E delay does not exceed the reliability constraint associated with its respective network. Subsequently, to account for the risk of loss incurred when reliability is not satisfied, the value-at-risk (VaR) concept defined as $\text{VaR}_{1-\alpha} = -\inf_{t \in \mathbb{R}}\{P(X \leq t) \geq 1 - \alpha\}$ [13], can be used. However, VaR is an uncoherent risk measure, making its analysis intractable. Thus, we define the EVaR as $\phi_t = \frac{\log \mathbb{E}[\exp(-\gamma Q_t)]}{\gamma}$, which is a coherent risk measure that corresponds to the tightest possible upper bound obtained from the VaR. In the EVaR, $Q_t := \max_{u \in \mathcal{U}}\{Q_{u,t}\}$ and $0 < \gamma \ll 1$. Subsequently, to ensure reliability, the following condition needs to be met: $\lim_{t \to \infty} \phi_t < \kappa$. Expanding the Maclaurin series of ϕ_t with respect to the log and exp functions we obtain, $\phi_t = E[Q_t] + \vartheta(Q_t) - 1 + \mathcal{O}(Q_t^3)$, where $\vartheta(Q_t) = E[Q_t^2] - E[Q_t]^2$ is the variance of the maximum queue length. Thus, to minimize $\lim_{t \to \infty} \phi_t$, it is sufficient to minimize the first two terms of its Maclaurin series. Consequently, we formulate the RIS association and phase shift-control problem for an RIS-assisted THz indoor network as follows:

$$\max_{\{\Pi_t\}} \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} x_{bu,t} \tilde{R}_{bu,t}, \quad (3)$$

$$\text{s.t.} \lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Q_t] < \varepsilon, \quad (4)$$

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[Q_t^2] < \eta, \quad (5)$$

$$\phi_{bu,t} \in \Phi, \forall b \in \mathcal{B}, \forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \quad (6)$$

$$\sum_{u \in \mathcal{U}} x_{bu,t} \leq 1, \forall b \in \mathcal{B}, \forall t \in \mathcal{T}, \quad (7)$$

$$x_{bu,t} \in \{0, 1\}, \forall b \in \mathcal{B}, \forall u \in \mathcal{U}, \forall t \in \mathcal{T}. \quad (8)$$

Here, maximizing the objective function in (3) ensures that the visual component of the VR experience is guaranteed, thus, delivering a seamless experience. On the other hand, (4) and (5) ensure that the constraint of mitigating the risk will be satisfied, where $\eta = \varepsilon^2 + 2[\gamma(\kappa + 1) - \varepsilon]$, this further guarantees that the haptic component of the VR will be delivered successfully. Given that the length of the queues changes with random events and their probability distribution is not known a priori, the optimization problem in (3) cannot be solved using traditional stochastic optimization techniques [16]. Next, we propose a tunable minimum-drift-plus-penalty optimization problem based on Lyapunov optimization to reformulate the problem stated previously.

B. Lyapunov Optimization

We use a Lyapunov optimization approach [16] to solve problem (3). This approach allows us to convert the constraints into a tractable form. Henceforth, to ensure (4) and (5), we define two virtual queues Z_1 and Z_2, having with the following dynamics:

$$Z_{1,(t+1)} = \max\{Z_{1,t} + Q_t - \varepsilon, 0\}, \quad (9)$$

$$Z_{2,(t+1)} = \max\{Z_{2,t} + Q_t^2 - \eta, 0\}. \quad (10)$$

Moreover, given that our initial optimization problem is a maximization problem, our aim is to minimize the drift-plus-penalty expression given by:

$$\Delta_t - V \sum_{b \in \mathcal{B}} \sum_{u \in \mathcal{U}} x_{bu,t} \tilde{R}_{bu,t}, \quad (11)$$

where $\Delta_t = \mathbb{E}[L_{t+1} - L_t|Q_t]$, L_t is the Lyapunov function given by $L_t = \frac{1}{2}(Z_{1,t}^2 + Z_{2,t}^2 + \sum_{u\in\mathcal{U}} Q_{u,t}^2)$. Next, we transform problem (3) into one whose objective is a linear weighted function, and its constraints are no longer a function of Q_t as in (4) and (5).

Proposition 1. *The conditional Lyapunov drift-plus-penalty bound under any feasible control policy π_t is formulated as follows:*

$$\Delta_t \leq \Upsilon + \sum_{u=1}^{U} Q_{u,t}(A_{u,t} - R_{bu,t}) + Z_{1,t}(Q_t - \varepsilon) + Z_{2,t}(Q_t^2 - \eta) - V \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}} x_{bu,t}\tilde{R}_{bu,t} \tag{12}$$

Proof: Given that for $\forall x \in \mathbb{R}, \max\{x, 0\}^2 \leq x^2$, we subtract $Q_{u,t}$ on both sides and square (2) as follows:

$$Q_{u,t+1}^2 - Q_{u,t}^2 \leq (Q_{u,t} - \tilde{R}_{bu,t})^2 + A_{u,t}^2 + 2A_{u,t}(Q_{u,t} - \tilde{R}_{bu,t}) - Q_{u,t}^2.$$

Simplifying the equation leads to the following:

$$\frac{Q_{u,t+1}^2 - Q_{u,t}^2}{2} \leq \frac{\left(\tilde{R}_{bu,t} - A_{u,t}\right)^2}{2} + Q_{u,t}(A_{u,t} - \tilde{R}_{bu,t})$$

Similarly,

$$\frac{Z_{1,t+1}^2 - Z_{1,t}^2}{2} \leq \frac{(Q_t - \varepsilon)^2}{2} + Z_{1,t}(Q_t - \varepsilon),$$
$$\frac{Z_{2,t+1}^2 - Z_{2,t}^2}{2} \leq \frac{\left(Q_t^2 - \eta\right)^2}{2} + Z_{2,t}(Q_t^2 - \eta).$$

After some mathematical manipulation, we obtain:

$$L_{t+1} - L_t \leq \Upsilon + \sum_{u=1}^{U} Q_{u,t}(A_{u,t} - \tilde{R}_{bu,t}) + Z_{1,t}(Q_t - \varepsilon) + Z_{2,t}(Q_t^2 - \eta). \tag{13}$$

where $\Upsilon = \frac{U\left(\max_{b\in\mathcal{B},u\in\mathcal{U},t} \tilde{R}_{bu,t}\right)^2 + \varepsilon^2 + \eta^2}{2}$. ∎

Thus, instead of minimizing the drift-plus-penalty expression, we minimize the maximum bound of the one-time slot conditional Lyapunov drift plus penalty in (12). The initial optimization problem is reformulated as:

$$\max_{\{\Pi_t\}} V \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}} x_{bu,t}\tilde{R}_{bu,t} - \sum_{u=1}^{U} Q_{u,t}(A_{u,t} - \tilde{R}_{bu,t}) - Z_{1,t}(Q_t - \varepsilon) - Z_{2,t}(Q_t^2 - \eta), \tag{14}$$

subject to

$$\phi_{bu,t} \in \Phi, \forall b \in \mathcal{B}, \forall u \in \mathcal{U}, \forall t \in \mathcal{T}, \tag{15}$$

$$\sum_{u\in\mathcal{U}} x_{bu,t} \leq 1, \forall b \in \mathcal{B}, \forall t \in \mathcal{T}, \tag{16}$$

$$x_{bu,t} \in \{0, 1\}, \forall b \in \mathcal{B}, \forall u \in \mathcal{U}, \forall t \in \mathcal{T}. \tag{17}$$

Thus, employing virtual queues and Lyapunov optimization allowed us to transform our optimization problem into a linear weighted function. Nevertheless, solving problem (14) using integer programming will be very complex due to its combinatorial nature and the stochasticity of the variables. Given that the distribution of system parameters is not characterizable, the problem cannot be solved using stochastic matching theory or stochastic optimization. Next, to solve (14) we propose a centralized and low-complexity RNN RL framework that provides the optimal policy for the RIS-UE association. Moreover, RNN RL is suitable for this problem since it can reduce the dimensionality of the large state space, while capturing dynamic temporal behaviors [17].

IV. Recurrent Neural Networks RL for Wireless VR in THz operated RIS network

In this section, an adaptive control policy based on a deep RL framework is proposed. The proposed framework will allow us to learn the policy to solve the problem of RIS-UE associations in (14). We model (14) as a Markov decision process (MDP) represented by the tuple $\{\mathcal{S}, \mathcal{A}, P, R\}$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, P is an unknown state transition function, $P(\boldsymbol{s}', \boldsymbol{s}, \boldsymbol{a}) = \Pr(\boldsymbol{s}_{t+1} = \boldsymbol{s}' \mid \boldsymbol{s}_t = \boldsymbol{s}, \boldsymbol{a}_t = \boldsymbol{a})$, and $R(\boldsymbol{a}_t, \boldsymbol{s}_t)$ is the reward function [18]. Our action space is the set of all possible RIS associations VR UEs, $\mathcal{A} = \{[x_{bu}]_{B\times u} \mid x_{bu} \in \{0, 1\}, b = 1, ..., B, u = 1, ..., U\}$ and the reward is $R(\boldsymbol{a}_t, \boldsymbol{s}_t) = V \sum_{b\in\mathcal{B}} \sum_{u\in\mathcal{U}} x_{bu,t}\tilde{R}_{bu,t} - \sum_{u=1}^{U} Q_{u,t}(A_{u,t} - R_{bu,t}) - Z_{1,t}(Q_t - \varepsilon) - Z_{2,t}(Q_t^2 - \eta)$ which is the current objective function in (14). The state is the set of VR UE queue lengths, virtual queue lengths, and the state of LoS links between RISs and VR UEs, $\mathcal{S} = \{[s_{bu}]_{B\times u}, [Q_u]_{U\times 1}, Z_1, Z_2, \mid s_{bu} \in \{0, 1\}, \{Q_u, Z_1, Z_2\} \in \{\mathbb{Z}^+, b = 1, ..., B, u = 1, ..., U\}$.

We represent the class of parameterized policies of our MDP as $\Pi_t = \{\pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t) \mid \theta \in \mathbb{R}^m\}$, where $\pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t) = \Pr\{\boldsymbol{a} = \boldsymbol{a}_t|\boldsymbol{s} = \boldsymbol{s}_t, \boldsymbol{\theta}\}$. The stochastic reward function $R(\boldsymbol{a}_t, \boldsymbol{s}_t)$ during next time slot has a transition probability of $\boldsymbol{\rho}_t = \prod_{\boldsymbol{s}'\in\mathcal{S}} \pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t) \Pr(\boldsymbol{s}_{t+1} = \boldsymbol{s}'|\boldsymbol{s}_t = \boldsymbol{s}, \boldsymbol{a}_t = \boldsymbol{a})$. To solve the optimization problem in (14), the controller needs to have full knowledge about the transition probability and all possible values of $R(\boldsymbol{a}_t, \boldsymbol{s}_t)$ for all possible states of MDP under a given policy $\pi_{\boldsymbol{\theta}}$. Given that our model is highly dynamic due the mobility of users and the nature of the channel, the transition probability of states cannot be characterized through probability distribution functions (PDFs). Thus, it is necessary to use an RL framework to solve (14). We particularly use a *policy-search* approach to find the optimal RIS to VR user association while maintaining a high reliability and a high data rate formulated in (14). For each policy, we define its value as:

$$J(\boldsymbol{\theta}) = \sum\nolimits_{\boldsymbol{s}'\in\mathcal{S}} R(\boldsymbol{a}_t, \boldsymbol{s}_t)\boldsymbol{\rho}_{\boldsymbol{t}}. \tag{18}$$

Hence, to find the optimal policy, we need to find $\theta^* = \arg\max_{\theta} J(\boldsymbol{\theta})$. To do so, we need to perform a gradient ascent on the policy parameters. Subsequently, we need to derive $\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta})$. Similarly to [18], by writing $\nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t) = \frac{\nabla_{\boldsymbol{\theta}} \pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t)}{\pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t)}$ we can reformulate (18), as follows:

$$\nabla_{\boldsymbol{\theta}} J(\boldsymbol{\theta}) \approx \mathbb{E}_{\Lambda_t}\{\nabla_{\boldsymbol{\theta}} \log \pi_{\boldsymbol{\theta}}(\boldsymbol{a}_t|\boldsymbol{s}_t) R(\boldsymbol{a}_t, \boldsymbol{s}_t)\}, \tag{19}$$

where $\Lambda_t = \{\boldsymbol{a}_t, \boldsymbol{s}_{t+1} = \boldsymbol{s}' \mid \boldsymbol{s}' \in \mathcal{S}\}$ is the trajectory of the MDP for the next time slot. Subsequently, we can use (19)

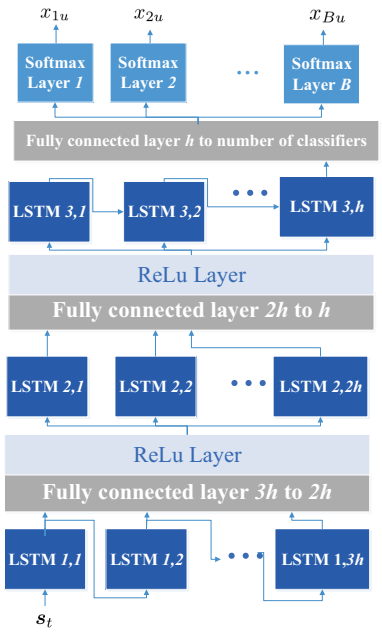

Fig. 2: Illustrative example of the RNN architecture.

to solve the optimization problem in (14) using a gradient ascent algorithm such as REINFORCE [18]. Nevertheless, given that the number of states is considerably high, this procedure will be intractable, which motivates the need for a function approximator through the use of DNNs. Given that the reliability depends on the prediction of the VR users mobility pattern that will determine their associations to RISs, it is important to implement a framework that is capable of capturing the dynamic behavior exhibited. To address these challenges, using an RNN to represent the policy of RL will extract the channel's dynamic features and learn an optimized sequence guaranteeing reliability at each time instant, based on the input features [17]. Since we deal with time-varying policies, it is natural to resort to RNNs. Indeed, RNNs are known to be effective in processing time-related data and capture dynamic temporal behaviors. As such, we represent the policy $\pi_{\boldsymbol{\theta}}(\boldsymbol{a}|\boldsymbol{s})$ by an RNN [19] that can learn the user associations to RISs. In particular, we use a many-to-many RNN as shown in Fig. 2. The overall RNN consists of an encoder and a decoder network: The encoder network comprises three long short-term memory (LSTM) layers, two fully connected layers and two rectified linear unit (ReLu) layers. The state of the MDP is thus encoded in the output of LSTM $3, h$. As for the decoder network, it consists of the last fully connected layer and the B softmax layers. Moreover, the input consists of the states $\boldsymbol{s}_t \in \mathcal{S}$ of the MDP that are fed to the first LSTM layer of h hidden layers. Subsequently, the B softmax layers output the actions of the MDP $a_t \in \mathcal{A}$, i.e., the bth softmax layer outputs $\{x_{bu} \mid x_{bu} \in 0, 1, b = 1, \ldots, B, u = 1, \ldots, U\}$. This architecture was chosen given that the LSTM layers allow us to avoid the problem of vanishing gradients [18]; more precisely compared to other DNNs, it provides a faster RL algorithm via

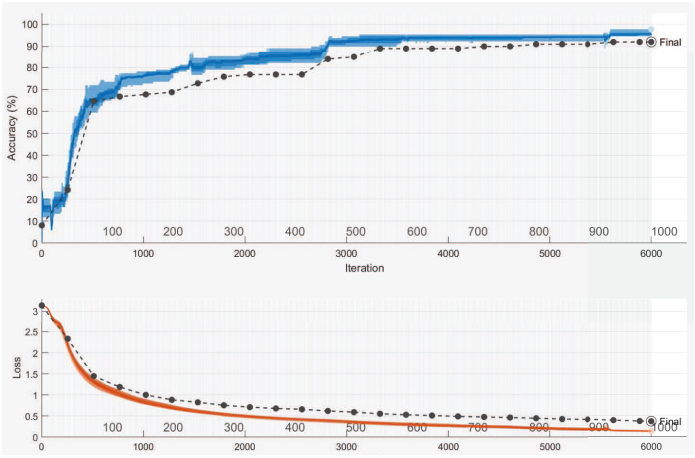

Fig. 3: The training and validation process of the RNN.

a slow RL. That is, instead of depending on the convergence speed of the RL algorithm through conventional gradient ascent, its policy is represented by an RNN. Subsequently, given that the RNN receives all the typical information that a regular RL algorithm would receive, the activations of the RNN store the state that would improve the speed of the RL algorithm on the current MDP [20].

V. Simulation Results and Analysis

For our simulations, we consider the following parameters: $T = 300\,\mathrm{K}$, $p = 1\,\mathrm{W}$, $M = 10\,\mathrm{Mbits}$, $f = 1\,\mathrm{THz}$, $W = 30\,\mathrm{GHz}$, $K(f) = 0.0016\,\mathrm{m}^{-1}$ with 1% of water vapor molecules as in [21]. The RISs are deployed over the 4 walls of an indoor area modeled as a square of size $40\,\mathrm{m} \times 40\,\mathrm{m}$. All statistical results are averaged over a large number of independent runs. In order to train the network, we consider the RNN architecture shown in Fig. 2 with a maximum epoch of 1000 and a minimum batch size of 128. Furthermore, the network was trained with data generated from VR users moving according to a *random walk* which constitutes the most general scheme characterizing users' mobility[2] [19]. In Fig. 3, we analyze the convergence of our proposed RNN-RL algorithm, in terms of accuracy and training loss. Note that, 80% of our dataset is used for the training process, 10% is used for the validation, and the remaining 10% is used for testing. In particular, the training process uses data generated from users' random walks and the optimal solution to fit the model. Subsequently, throughout the validation process, the random walks' data is used to provide unbiased estimators of the hyper-parameters corresponding to the RNN. Fig. 3 shows a validation accuracy of 91.92%. Both the accuracy and loss of the training and validation processes show smoothness in the curve. Finally, after obtaining all the hyper-parameters of the RNN, the test dataset can provide an unbiased evaluation of a final model fit. As such, simulation results show a testing error of 0.97%.

[2]Our approach can accommodate any other mobility model.

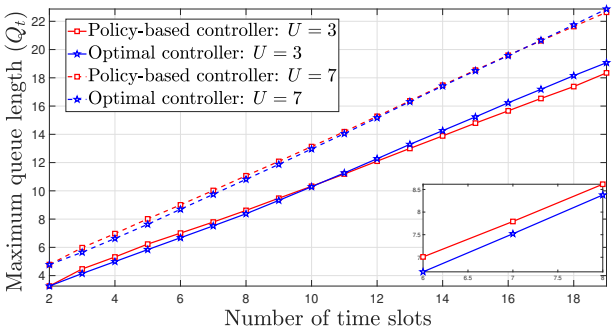

Fig. 4: Maximum queue length vs. number of time slots.

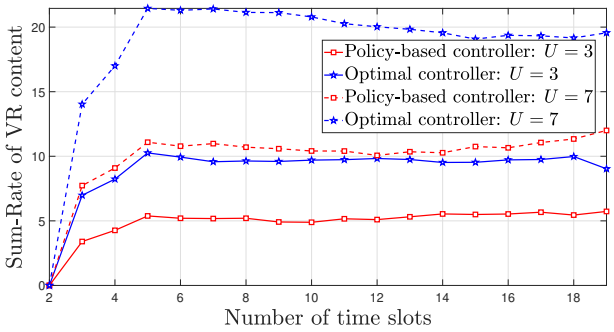

Fig. 5: Sum-rate of VR content vs. number of time slots.

We compare our RNN RL policy to the optimal solution for different number of users. In Fig. 4, the maximum queue length over the number of iterations is plotted. Given that the RNN was capable of capturing the dynamic behavior of the channel, for $U = 7$, the maximum queue length for the policy-approach is only 0.52% higher than the optimal maximum queue length. Meanwhile, for $U = 3$ it is nearly equal to the optimal solution. Thus, this confirms that combining the risk-based approach with the RNN RL leads to highly reliable results. Clearly, the maximum queue length increases as the number of VR users in the network increases. Fig. 5 shows the sum-rate of VR content, which constitutes the objective function in (14), over time. Here, the policy-based controller offers a solution that is considerably farther than the optimal solution in comparison to the results obtained for the maximum queue length in Fig. 4. The reason for this is that a sum-term rate is being compared rather than individual rates; thus, the inaccuracy in every measure propagates into a higher inaccuracy when terms are added. As we can see, for $U = 7$, the sum-rate for the policy-approach is 22% less than optimal sum-rate. As for $U = 3$, it is 35% less than the optimal solution. Clearly, as the number of VR users increases, the sum-rate increases, and so does the gap between the optimal solution and the policy approach.

VI. Conclusion

In this paper, we have investigated the problem of RIS-VR user association while guaranteeing reliable, low latency and high rate communications. We have proposed a risk-based aware optimization problem that takes into account the higher order statistics of the queue length, thus guaranteeing continuous reliability. The proposed problem was further transformed using Lyapunov optimization to a linear weighted function. Furthermore, the problem was solved using an RNN RL framework to reduce the dimensionality of the state space and capture channel dynamics and user mobility.

References

[1] M. Chen, W. Saad, and C. Yin, "Virtual reality over wireless networks: Quality-of-service model and learning-based resource management," *IEEE Transactions on Communications*, vol. 66, no. 11, pp. 5621–5635, Jun. 2018.

[2] W. Saad, M. Bennis, and M. Chen, "A vision of 6G wireless systems: Applications, trends, technologies, and open research problems," *arXiv preprint arXiv:1902.10265*, 2019.

[3] G. Gerardino, B. Alvarez, K. Pedersen, and P. Mogensen, "MAC layer enhancements for ultra-reliable low-latency communications in cellular networks," in *Proc. of IEEE International Conference on Communications (ICC)*, Paris, France, pp. 1–7.

[4] M. Steeg and A. Stöhr, "High data rate 6 gbit/s steerable multibeam 60 ghz antennas for 5g hot-spot use cases," in *Proc. of IEEE Photonics Society Summer Topical Meeting Series (SUM)*, San Juan, Puerto Rico, Jul. 2017, pp. 141–142.

[5] A. Moldovan, P. Karunakaran, I. F. Akyildiz, and W. H. Gerstacker, "Coverage and achievable rate analysis for indoor terahertz wireless networks," in *Proc. of IEEE International Conference on Communications (ICC)*, Paris, France, Jul. 2017, pp. 1–7.

[6] E. Basar, M. Di Renzo, J. de Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *arXiv preprint arXiv:1906.09490*, 2019.

[7] A. T. Z. Kasgari, W. Saad, and M. Debbah, "Human-in-the-loop wireless communications: Machine learning and brain-aware resource management," *to appear, IEEE Transactions on Communications*, 2019.

[8] M. Chen, W. Saad, and C. Yin, "Echo-liquid state deep learning for 360 content transmission and caching in wireless vr networks with cellular-connected UAVs," *IEEE Transactions on Communications*, vol. 67, no. 9, pp. 6386–6400, 2019.

[9] J. Park and M. Bennis, "URLLC-eMBB slicing to support VR multimodal perceptions over wireless cellular systems," in *Proc. of IEEE Global Communications Conference (GLOBECOM)*, Abu Dhabi, UAE, Dec. 2018.

[10] A. S. Cacciapuoti, R. Subramanian, K. R. Chowdhury, and M. Caleffi, "Software-defined network controlled switching between millimeter wave and terahertz small cells," *arXiv preprint arXiv:1702.02775*, 2017.

[11] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[12] M. Jung, W. Saad, Y. Jang, G. Kong, and S. Choi, "Uplink data rate in large intelligent surfaces: Asymptotic analysis under channel estimation errors," in *submitted to Proc. of IEEE Wireless Communications and Networking Conference (WCNC), Marrakech, Morocco*, 2019.

[13] A. Ahmadi-Javid, "Entropic value-at-risk: A new coherent risk measure," *Journal of Optimization Theory and Applications*, vol. 155, no. 3, pp. 1105–1123, 2012.

[14] C. Chaccour, R. Amer, B. Zhou, and W. Saad, "On the reliability of wireless virtual reality at terahertz (THz) frequencies," in *Proc. of the 10th IFIP International Conference on New Technologies, Mobility and Security (NTMS)*, Canary Islands, Spain, June. 2019.

[15] R. Zhang, K. Yang, Q. H. Abbasi, K. A. Qaraqe, and A. Alomainy, "Analytical modelling of the effect of noise on the terahertz in-vivo communication channel for body-centric nano-networks," *Nano communication networks*, vol. 15, pp. 59–68, 2018.

[16] M. J. Neely, "Stochastic network optimization with application to communication and queueing systems," *Synthesis Lectures on Communication Networks*, vol. 3, no. 1, pp. 1–211, 2010.

[17] M. Chen, U. Challita, W. Saad, C. Yin, and M. Debbah, "Artificial neural networks-based machine learning for wireless networks: A tutorial," *IEEE Communications Surveys & Tutorials*, 2019.

[18] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*. MIT press, 2018.

[19] M. Naderi Soorki, W. Saad, and M. Bennis, "Ultra-reliable millimeter-wave communications using an artificial intelligence-powered reflector," in *Proc. IEEE Global Communications Conference*, Waikoloa, HI, USA, Dec. 2019.

[20] Y. Duan, J. Schulman, X. Chen, P. L. Bartlett, I. Sutskever, and P. Abbeel, "Rl2: Fast reinforcement learning via slow reinforcement learning," *arXiv preprint arXiv:1611.02779*, 2016.

[21] C.-C. Wang, X.-W. Yao, C. Han, and W.-L. Wang, "Interference and coverage analysis for terahertz band communication in nanonetworks," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Singapore, Singapore, Dec. 2017.