Learning-based mmWave V2I Environment Augmentation through Tunable Reflectors

Lan Zhang*, Xianhao Chen*, Yuguang Fang*, Xiaoxia Huang†, Xuming Fang‡
* Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611, USA
† School of Electronics and Communication Engineering, Sun Yat-Sen University, Guangzhou, 510275, China
‡ Key Lab of Information Coding and Transmission, Southwest Jiaotong University, Chengdu, 610031, China
E-mail: lanzhang@ufl.edu; xianhaochen@ufl.edu; fang@ece.ufl.edu; xiaoxiah@gmail.com; xmfang@swjtu.edu.cn

***Abstract*—To support the demand of multi-Gbps sensory data exchanges for enhancing (semi)-autonomous driving, millimeter-wave bands (mmWave) vehicular-to-infrastructure (V2I) communications have attracted intensive attention. Unfortunately, the vulnerability to blockages over mmWave bands poses significant design challenges, which can be hardly addressed by manipulating end transceivers, such as beamforming techniques. In this paper, we propose to enhance mmWave V2I communications by augmenting the transmission environments through reflection, where highly-reflective cheap metallic plates are deployed as tunable reflectors without damaging the aesthetic nature of the environments. In this way, alternative indirect line-of-sight (LOS) links are established by adjusting the angle of reflectors. Our fundamental challenge is to adapt the time-consuming reflector angle tuning to the highly dynamic vehicular environment. By using deep reinforcement learning, we propose the Learning-based Fast Reflection (LFR) algorithm, which autonomously learns from the observable traffic pattern to select desirable reflector angles in advance for probably blocked vehicles in near future. Simulation results demonstrate our proposal could effectively augment mmWave V2I transmission environments with significant performance gain.**

***Index Terms*—Blockages, learning-based, mmWave, tunable Reflector, transmission environment augmentation, V2I.**

I. Introduction

Hundreds of sensors are equipped to vehicles nowadays for enhancing driving safety and convenience, generating a large amount of sensing data [1]. Vehicular-to-infrastructure (V2I) communications allow sharing these sensing data among vehicles and roadside units (RSUs), which enable more advanced services, ranging from safety-related forward collision warning to future automatic driving [2]. To support such large volume data exchanges, millimeter wave (mmWave) bands have attracted great attentions [1], [2]. However, in the highly dynamic vehicular environments, the poor transmission characteristics of mmWave bands, especially the vulnerability to blockages, become more severe [2], [3]. Due to the poor penetration and diffraction, as well as the nature of directional transmission, an mmWave V2I link can be easily blocked by obstacles as shown in Fig.1(a), which can not be solved by increasing the transmit power or antenna gain.

Recently, alternative line-of-sight (LOS) links have been considered to address the blocked mmWave transmissions [4]–[8]. One intuitive scheme is to utilize mmWave relays either on moving vehicles or statically deployed along the roadside.

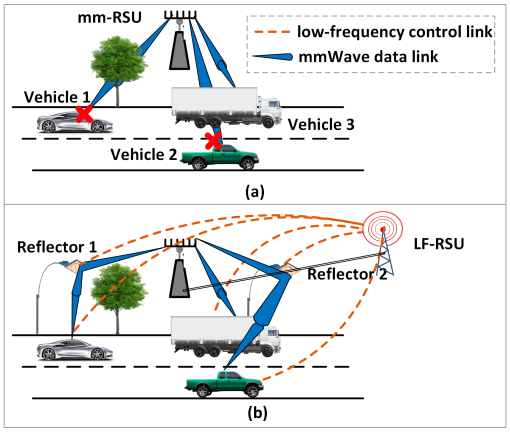

Fig. 1. The illustration of blockages in mmWave V2I communications (a); the system architecture for tunable reflector enabled transmission environment augmentation (b).

However, on the one hand, due to the limited height of a vehicle, the onboard relays usually have very short LOS transmission range [3], [9]. Thus, a blocked V2I transmission needs to be recovered by cooperation among multiple onboard relays, which introduces a large overhead in terms of delay, energy and spectrum utilization, etc. Although a roadside relay can be deployed at a relatively high position with longer LOS transmissions, a large-scale deployment of these expensive mmWave relays will seriously burden the cost of mmWave V2I communications [9]. Instead of using expensive mmWave relays, alternative LOS links established through reflection have aroused great attention recently [4]–[8]. Through bouncing off surrounding environments like wall and floor, Xue et al. proposed to concurrently transmit multiple reflected beams to one user [4]. However, the poor reflectance of surrounding environments like concrete and wood usually leads to severe signal loss [6]. Fortunately, the cheap metallic plates with smooth surface can reflect nearly 100% signal strength [10]. Using metallic ceiling, Zhou, et al. enabled several machine pairs to simultaneously communicate through mmWave bands in a data center, whose effectiveness is verified through the testbed [5]. Unlike data centers used to store machines, our previous work considered the functionality and aesthetic nature of indoor human living environment, and proposed to deploy small-piece reflectors to augment the transmis-

sion environments of mmWave wireless local area networks (WLANs) [8]. Moreover, to deal with the limited augmentation caused by the strict reflection requirements (identical incidence and reflection angle), we proposed to utilize tunable reflectors, i.e., highly reflective cheap metallic plates with adjustable angles, to adaptively change the directions of the reflected beams, whose effectiveness is theoretically verified [8]. Although tunable reflectors can promisingly enhance mmWave transmissions by introducing alternative LOS links in a cost-effective way, our previous work designed for WLANs can be hardly implemented to highly dynamic V2I communications. To the best of our knowledge, none of existing analysis has focused on this important design.

In this paper, we investigate how to implement tunable reflectors to augment transmission environments for better mmWave V2I communications. Without damaging the aesthetic and functionality nature of driving environments, the small-piece tunable reflectors can be deployed on the top of roadside lighting poles. To effectively tune the angle of a reflector plate, we first design the system architecture followed by the operational procedures. Considering the contradiction between the highly dynamic blockages and the time-consuming mechanical angle tuning, our fundamental challenge is to design the angle tuning policy to fast adapt to V2I transmissions. To enable autonomously exploration, learning and mapping from the observable traffic patterns to desirable reflector angles, we propose the Learning-based Fast Reflection (LFR) algorithm by using deep reinforcement learning (DRL). Embracing the advantages of deep neural networks (DNN), DRL has been adopted in complicated vehicular transmission environments [11], [12]. By using LFR algorithm, the angle of a reflector plate can be determined to serve the blocked vehicle in near future based on current traffic pattern, where a reflective link can be established in advance to recover the blocked data to the fullest extent. Simulation results verify the effectiveness of our environment augmentation proposal for mmWave V2I communications, and the significant performance gain by LFR algorithm.

II. System Model

A. System Architecture

Our system architecture is extended from our previous work for mmWave wireless local area networks (WLANs) [8]. As illustrated in Fig.1(b), the system architecture is developed form the concept of control-/data-plane decoupling to mitigate the inherent blindness of directional mmWave transmissions while saving cost. The data-plane provides mmWave V2I transmissions among the RSUs using mmWave bands, named by mm-RSUs, and vehicles. The control-plan aims at coordinating the establishments of reflective links by supporting fast handshakes among the vehicle, reflector(s) and mm-RSU, and facilitating operations on directional mmWave transmissions, which is done by a controller. Physically, the controller could be a RSU using relatively reliable low-frequency bands with omni-directional coverage, named by LF-RSU. Based on the control signaling, the angle of a reflector plate is adjusted to reflect mmWave signals to surrounding vehicles. Besides a smooth metallic plate, a tunable reflector is equipped with a cost-effective low-frequency transceiver, such as the commercialized onboard V2I transceiver, and a simple mechanical device to rotate the metallic plate. Without damaging the aesthetic nature of environments, these small-piece tunable reflectors are deployed at places of interest, such as the top of a lighting pole. To enable a reflective V2I link, the controller connects with an mm-RSU through either low-frequency links or deployed wired links. A vehicle is assumed to be equipped with both the low-frequency and mmWave transceivers for control-/data-plane signaling, respectively.

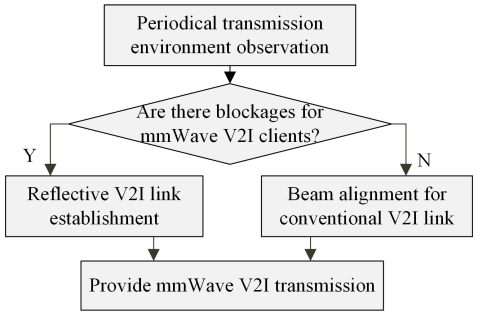

Fig. 2. The operational procedures of our tunable reflector enabled mmWave V2I communications.

B. Operational Procedures

Based on above system architecture, we propose the corresponding operational procedures as illustrated in Fig.2. In each reflector scheduling slot $t \in \mathcal{T}$, instead of the conventional V2I beam alignment which might be failed due to blockages as shown in Fig.1(a), we first observe the environment to examine blockages. Since the blockages can be either dynamic like moving vehicles or static like trees or buildings, the examination is based on the information of both current traffic pattern and existing environment information (historical blockage records). After that, the LOS V2I links can be established based on conventional beamforming techniques, while the blocked V2I links would be assisted by tunable reflectors. By using either conventional or reflective V2I links, the mmWave transmissions can be processed. It should be mentioned that the time-scale of blockage examinations is larger than that of communication scheduling, e.g., hundreds or thousands vs. tens of milliseconds. The duration of reflector scheduling can sufficiently catch the traffic changing, i.e., dynamic blockages in mmWave V2I transmissions, and meet the time requirement for mechanical reflector angle tuning. Based on the LOS V2I tunnels provided by tunable reflectors in one reflector scheduling slot, a vehicle may have multiple data transmissions, depending on the communication scheduling for specific applications.

The transmission environment observation focuses on the road segment covered by a typical mm-RSU, which is assumed to be two-way l-lane. The driving information of all vehicles, i.e., vehicles on or heading to this road segment, are recorded and periodically updated in the LF-RSU through the

control signaling. This kind of driving data collection might be a mandate for driving-safety related applications, such as collision warning and traffic management in near future [1]. Denote all vehicles at time t by a set $\mathcal{V}_t = \{v_{t,i}\}_{i=1,\dots,V_t}$ of V_t vehicles, whose driving information can be given by $I^v = \{x^v, y^v, d, u, h\}$, representing the coordinate of current position, heading direction, velocity and size, respectively. Since there might be only a portion of all vehicles subscribe mmWave V2I services, denote vehicles with mmWave V2I requests by a set $\mathcal{V}_t^m = \left\{v_{t,i}^m\right\}_{i=1,\dots,V_t^m} \subset \mathcal{V}_t$ of V_t^m vehicles. In addition, the existing environment information of this road segment illustrates the static blockages, which is mapped to the definitely blocked positions based on the historical knowledge. Denote the existing environment information by $\mathcal{I}^e = (\mathcal{X}^b, \mathcal{Y}^b)$, which might require an updated with a large-scale time duration.

Based on the above observations, at each time slot $t \in \mathcal{T}$, tunable reflectors are activated to serve the blocked mmWave V2I clients. Assume each reflector is only used by one mm-RSU. Define a set $\mathcal{K} = \{k_j\}_{j=1,\dots,K}$ of K tunable reflectors, deployed along the roadside at the top of the lighting poles with height h^r, whose positions are $(x_{k_j}^r, y_{k_j}^r)$. Since the mechanical angle tuning is time-consuming, we prefer to use the reflector with a small angle tuning. Denote the angle of reflectors $k_j \in \mathcal{K}$ at time t by ϕ_{t,k_j}. Thus, we have the status information of tunable reflectors, $I_{t,k_j}^r = \left\{x_{k_j}^r, y_{k_j}^r, \phi_{t,k_j}\right\}_{k_j \in \mathcal{K}}$, which is recorded by the LF-RSU. By integrating all above contextual information, we particularly design the learning-based fast reflection (LFR) algorithm to autonomously handle the whole operational procedures for better mmWave V2I communications, detailed below.

III. LFR Algorithm

A. The Learning Framework

Aiming at autonomously handling transmission environment augmentation for blocked mmWave V2I transmissions, LFR algorithm is based on deep reinforcement learning (DRL). As a policy learning process, the agent of DRL periodically makes decisions, observes and learns from the corresponding results, and then autonomously adjust its policy [11]–[13]. The tasks are usually formulated as Markov Decision Process (MDP) with implicit transition probability and the rewarding. Embracing the advantages of deep neural networks (DNN), DRL can be fast trained to reach optimal policy for complicated tasks, which has been applied in recent mmWave vehicular analysis [11]–[13]. As illustrated in Fig.3, our agent periodically interacts with the environment to make decisions about mmWave V2I transmissions. Physically, the agent can be an mm-RSU, whose observations are gathered from the LF-RSU through wired connections.

At each time slot $t \in \mathcal{T}$, the agent observes a state s_t from state space $\mathcal{S}$, which includes the traffic pattern $\mathcal{I}^v = \left\{I_{v_i}^v | v_i \in \mathcal{V}\right\}$, the existing environment $\mathcal{I}^e = (\mathcal{X}^b, \mathcal{Y}^b)$,

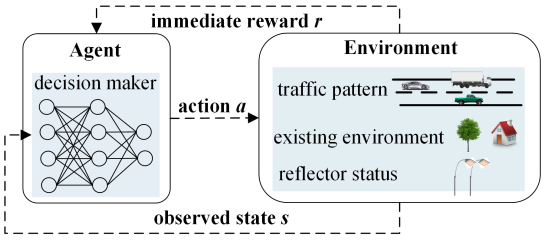

Fig. 3. The framework of LFR algorithm.

the reflector status information $\mathcal{I}^r = \left\{I_{k_j}^r | _{k_j \in \mathcal{K}}\right\}$, and the mmWave V2I client $\mathcal{V}^m$. Thus, we have

$$s_t = \{\mathcal{I}_t^r, \mathcal{I}_t^v, \mathcal{I}^e, \mathcal{V}_t^m\}. \tag{1}$$

Based on state s_t, the agent takes an action $a_t \in \mathcal{A}$ to determine the establishment of reflective links, i.e., a reflector should be used at what specific angle for which blocked vehicle. Define the blocked mmWave V2I clients in time slot t by a set $\mathcal{V}_t^{bm} \subset \mathcal{V}_t^m$ of V_t^{bm} vehicles. Note that due to the limited number of tunable reflectors, given each reflector only serves one vehicle at one time, only part of blocked vehicles can be served when $V_t^{bm} > K$. Define the reflector allocation indicator by $J_{v_{t,i}^{bm}}^{k_j}$ with $J_{v_{t,i}^{bm}}^{k_j} = 1$ if the blocked vehicle $v_{t,i}^{bm} \in \mathcal{V}_t^{bm}$ is served by reflector $k_j \in \mathcal{K}$, and $J_{v_{t,i}^{bm}}^{k_j} = 0$ otherwise. Since one reflector can serve one vehicle at a time slot, $\sum_{v_{t,i}^{bm} \in \left\{\mathcal{V}_t^{bm} \cup v_{t,0}^{bm}\right\}} J_{V_{t,i}^{bm}}^{k_j} = 1$ for all $k_j \in \mathcal{K}$, where $v_{t,0}^{bm}$ represents reflector k_j is not used for any vehicle. Meanwhile, since we assume one blocked vehicle is at most served by one reflector, $\sum_{k_j \in \mathcal{K}} J_{v_{t,i}^{bm}}^{k_j} \in \{0, 1\}$ for all $v_{t,i}^{bm} \in \mathcal{V}_t^{bm}$. Due to certain beam-width, e.g., 15°, instead of high-resolution control, we divide the angle tuning range into M partitions to reduce the requirement of the mechanical tuning device while guaranteeing the performance of the reflected beams. At time slot t, the target angel of reflector $k_j \in \mathcal{K}$ should be determined by the service indicator $\left\{J_{v_{t,i}^{bm}}^{k_j}\right\}_{v_{t,i}^{bm} \in \mathcal{V}_t^{bm}}$. The action $a_t \in \mathcal{A}$ can be given by

$$a_t = \left\{J_{v_{t,i}^{bm}}^{k_j} | k_j \in \mathcal{K}, v_{t,i}^{bm} \in \mathcal{V}_t^{bm}\right\}. \tag{2}$$

Since $V_t^{bm} \leq V_t^m$, the dimension of action space is $V_t^m \times K$.

B. Expected Reward Estimation

The objective of LFR algorithm is to effectively serve the blocked mmWave V2I transmissions to mitigate their data loss at the maximum extent. On the one hand, compared with the blocked vehicles that can be fast recovered, vehicles with longer blocking time, i.e., more data loss, should have higher service priority. On the other hand, less angle tuning time should be used to leave more time for data transmission, and thus improve the utility of each reflective link. Since the learning process is guided by the reward value r, the reward function is defined as the increased data amount by introducing

reflective links. Denote the capacity of an mmWave V2I transmission for vehicle $v_{t,i}^m$ by

$$C_{v_{t,i}^m}(L(d)) = W\log_2(1+\frac{P_tG_0L(d)^{-1}}{\sigma^2}), \tag{3}$$

where P_t is the transmit power, G_0 is the beamforming gain, σ^2 is the noise power. $L(d) = \alpha + 10\eta\log_{10}(d) + \xi[dB], \xi \sim N(0,\rho^2)$ is the path loss model of mmWave transmissions, where d is the transmission distance, α and η are the least square fits of floating intercept and slope over the measured distance, and ρ^2 is the lognormal shadowing. The values of α, η and ρ are different for LOS and Non-LOS (NLOS, i.e., blocked) link states [14]. Thus, the capacity improvement by introducing a reflective link can be given by $\Delta C_{v_{t,i}^{bm}} = C^R_{v_{t,i}^{bm}}(L'(d')) - C^D_{v_{t,i}^{bm}}(L(d))$, $v_{t,i}^{bm} \in \mathcal{V}_t^{bm}$.

Since the mechanical angle tuning is time-consuming, we take the useful transmission time in each scheduling slot $t \in \mathcal{T}$ into account. Given the M partitioned angle tuning range, define the time used to tune to the next angle partition by $\Delta\tau$. Thus, the time used to tune reflector k_j from current angel $\phi_t^{k_j}$ to serve vehicle $v_{t,i}^{bm}$ can be expressed as $\tau_t^{k_j}(\phi_t^{k_j}, v_{t,i}^{bm})$,where $\tau_t^{k_j} \in \{m\Delta\tau\}_{m=0,1,\ldots,M}$. Given the duration of a time slot by Δt, we derive the reward function at time slot t by,

$$r_t = \sum_{k_j\in\mathcal{K}}\left(\sum_{v_{t,i}^{bm}\in\mathcal{V}_t^{bm}} J_{v_{t,i}^{bm}}^{k_j}\Delta C_{v_{t,i}^{bm}}\right)\left(\Delta t - \tau_t^{k_j}\right). \tag{4}$$

In order to obtain the long-term benefits, not only the immediate rewards but also the future rewards should be considered. Therefore, the aim of our agent is to find an optimal policy $a_t^* = \pi^*(x_t) \in \mathcal{A}$ for $t \in \mathcal{T}$, which maximizes the expected cumulative discounted rewards,

$$R_t = \mathbb{E}\left[\sum_{n=0}^{\infty}\beta^n r_{t+n}\right], \tag{5}$$

where $\beta \in [0,1]$ is the discount factor to limit the effect of rewards in the far future.

C. Deep Q-Learning

Based on a policy π, the agent maps the state space, $\mathcal{S}$, to the action space, $\mathcal{A}$, by $\pi : s_t \in \mathcal{S} \rightarrow a_t \in \mathcal{A}$, to maximize the long-term expected cumulative discount rewards, R_t. Q-learning algorithms are popular techniques to derive the optimal policy π^* due to its effectiveness and simplicity [11]–[13]. The Q-value $Q(s_t, a_t)$ is defined as the expected cumulative discounted reward, R_t, for a given state-action pair (s_t, a_t), which measures the qualify of a certain action a_t for a given state s_t. Based on the Q-value of a given state s_t, an improved policy can be easily derived by taking the action $a_t = \arg\max_{a\in\mathcal{A}} Q(s_t, a)$. Therefore, the optimal policy can be known based on the following update equation,

$$\begin{aligned} Q_{\text{new}}(s_t,a_t) =& Q_{\text{old}}(s_t,a_t) + \gamma[r_{t+1} + \\ & \delta\max_{s\in\mathcal{S}} Q_{\text{old}}(s_t,a_t) - Q_{\text{old}}(s_t,a_t)], \end{aligned} \tag{6}$$

where γ is the learning rate, and the second term on the right is the temporal difference error of the Q-value update, which approaches to zero under the optimal policy π^*. It has been proved that the Q-value in MDP problems converges to the optimum, if each action in $\mathcal{A}$ can be executed under each state in $\mathcal{S}$ for infinite times with appropriately decaying learning rate γ [11]–[13]. The optimal policy π^* can be determined based on the optimal Q-value Q^*. In classical Q-learning, the optimal policy is found through a look-up table of Q-value based on the state-action space, whose efficiency is limited by the size of that space. When the state-action space is huge, the Q-value of the infrequently visited state-action pair will be updated rarely, and thus the Q-value becomes hardly to converge. Recently, embracing the advantages of DNN, deep Q-learning algorithms have been widely adopted to handle MDP problems under a huge state-action space, where the Q-function is approximated by a DNN with weights $\boldsymbol{\theta}$ that is denoted as Q-network $\mathcal{Q}$ [11]–[13]. Based on the well-trained Q-network $\mathcal{Q}^*$, the sophisticated mapping between the state and action spaces can be used to determine the optimal Q-values, i.e., the optimal policy π^*. Given the training data set $\mathcal{D}$ with transitions $\{s_t, a_t, r_t, s_{t+1}\}_{t\in\mathcal{T}}$, the Q-network $\mathcal{Q}$ is trained based on the loss function below,

$$L(\boldsymbol{\theta}) = \sum_{(s_t,a_t)\in\mathcal{D}}(y - Q(s_t,a_t,\boldsymbol{\theta}))^2, \tag{7}$$

where $y = r_t + \max_{a\in\mathcal{A}} Q_{\text{old}}(s_t, a, \boldsymbol{\theta})$.

Algorithm 1: Pseudocode of LFR Algorithm

Data: $\mathcal{D} = \{s_t, s_{t+1}, a_t, r_t\}_{t\in\mathcal{T}}$, Q-network structure;
Result: Q-network $\boldsymbol{\theta}$;
Initialization: Q-network $\mathcal{Q}$ with random weights $\boldsymbol{\theta}$,
  experience memory $\mathcal{D}_M$, $episode = 1$;
while $episode < N_{epi}$ **do**
  With probability ϵ select a random action a_t, otherwise select $a_t = \arg\max_a Q(s_t, a; \boldsymbol{\theta})$;
  Perform action a_t and observe the reward r_t and next state s_{t+1}, and then store the new transaction (s_t, a_t, r_t, s_{t+1}) in D_M;
  Sample a minibatch of transitions from D_M;
  Optimize the Q-network $\mathcal{Q}$ by using this minibatch transactions through gradient descent with respect to weights $\boldsymbol{\theta}$ to minimize $L(\boldsymbol{\theta})$;
  $episode = episode + 1$;
end

D. LFR Algorithm Description

Based on above design, the proposed deep Q-network is trained on a data set $\mathcal{D}$, generated from the interactions with an environment simulator, Simulation of Urban MObility (SUMO) [15]. As an open source microscopic traffic simulator, SUMO enables repeatable controlled experiments at a per-vehicle level based on real-world traffic topology [15]. Specifically, we utilize SUMO to generate the traffic pattern

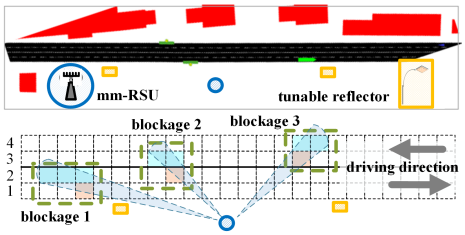

Fig. 4. The simulation scenario and dynamic blockages.

TABLE I
DEFAULT PARAMETERS OF EVALUATION

Parameter	Value	Description
α_L,η_L,α_N,η_N	72,2.92,61.4,2	channel state parameters
W	500MHz	bandwidth
P_t / σ^2	5×10^5	mm-RSU Tx/noise power
G_0	10dB	antenna gain
h^v	1.75m/3.5m	height of large/small vehicle
h^r	4.5m	height of a reflector
Δt	1s	scheduling interval

$I^v = \{x^v, y^v, d, u, s\}$, i.e., the position, direction, velocity, and shape of each vehicle on a road segment, and derive the existing environment information $\mathcal{I}^e$ based on the roadside environments. Combining with the initially randomized reflector status information $\mathcal{I}^r$, the immediate reward r_t of each action $a_t \in \mathcal{A}$ can be derived.

The detailed LFR algorithm is illustrated in Algorithm 1. During the training process, we utilize the Q-learning with experience replay technique [11]–[13] to improve the training performance by suppressing the temporal correlation of transactions in $\mathcal{D}$. Specifically, in each episode, a minibatch transactions is randomly sampled form the experience memory $\mathcal{D}_M$, which is used to train the Q-network $\mathcal{Q}$ through gradient descent method. Note that the new transactions will pop up the old ones in the experience memory $\mathcal{D}_M$, when the buffer of $\mathcal{D}_M$ is full. In addition, the ϵ-greedy policy is utilized to balance the exploration and exploitation, i.e., the trade-off between improving the system knowledge about the reward distribution (exploration) and switching to the action with the highest empirical mean reward (exploitation).

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of our LFR algorithm. As shown in Fig.4, a two-way 4-lane road segment is generated by SUMO simulator [15], where one mm-RSU and two tunable reflectors are uniformly deployed along the roadside. For simplicity, we evaluate two type vehicles with different sizes, e.g., sedan and bus/truck, where a large vehicle may block a small one in the adjacent lane under scenarios as shown in Fig.4. Based on the floating car data (FCD) output in SUMO, the driving information of every vehicle along this road can be collected in each time slot, including the position, velocity and corresponding size, i.e., $\mathcal{I}_t^v$. The default system parameters are listed in Table I, where subscripts 'L' and 'N' represent LOS and NLOS conditions, respectively [12], [14]. In addition, the Q-learning network in our simulation is a five-layer fully connected neural network with three hidden layer with 256, 128 and 128 neurons, respectively. The Relu function is used, and the learning rate γ is 0.01 at beginning and decreases exponentially.

To evaluate the performance of our augmented V2I transmissions, we first analyze the data loss caused by blockages, which involves only dynamic blockages, i.e., outages caused by large vehicles that are much more challenging comparing with static blockages. The total arrival rate of this road segment (4 lanes) is set to be 0.6. Since the large vehicles,

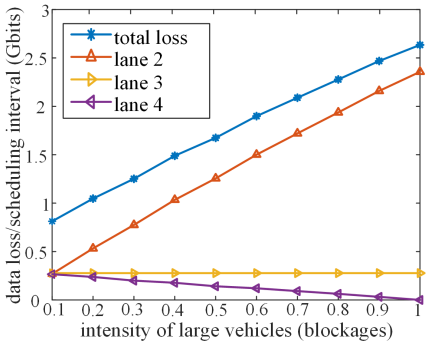

Fig. 5. The data loss caused by dynamic blockages.

like buses in the cities and trucks on the highway, are asymmetrically distributed among different lanes, i.e., they usually stay on the right-hand lane with relatively slow velocity, we dynamically change the intensity of large vehicles on the right-hand lanes (lane 1 and 4), where the intensity of large vehicles on the left-hand lanes (lane 2 and 3) are set to be 0.1. As illustrated in Fig.5, the average data loss in each scheduling interval is as a function of the intensity of large vehicles on the right-hand lanes. We observe that the data loss increases with the intensity of large vehicles. Even with a few large vehicles, the data loss is still significant due to the high transmission rate of mmWave transmissions, where the minimum total data loss is around 0.81 Gigabits per second. Thus, mitigation strategies are imperatively necessary for mmWave V2I communications. In addition, we observe that vehicles in the right-hand lanes (lane 2 and 3) have larger data loss compared with those in the left-hand lane (lane 4), which further motivates our analysis to provide fairness to vehicles with different sizes and driving on different lanes.

Based on the fact of poor system performance caused by blockages, we then evaluate the performance benefits through reflections as a function of the intensity of large vehicles. As illustrated in Fig.6, we observe that the ideal reflection, i.e., reflective links established without reflector angel tuning, can mitigate more than 60% data loss. However, due to the time-consuming nature of mechanical angle tuning, it is hard to eliminate the overhead of reflector angle tuning which is further analyzed below. Assume the unit time of angle tuning $\Delta\tau$ is 10 ms. Although comparing with the ideal reflection, the recovered data based on LFR scheme is reduced by nearly 8%, LFR scheme can still significantly recover more than 55% blocked data. Moreover, we observe that LFR scheme can effectively handle large data loss caused by more blockages,

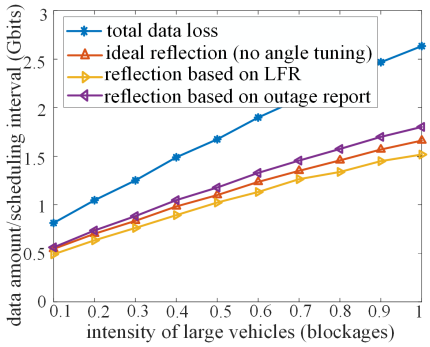

Fig. 6. Comparison of recovered data through reflections.

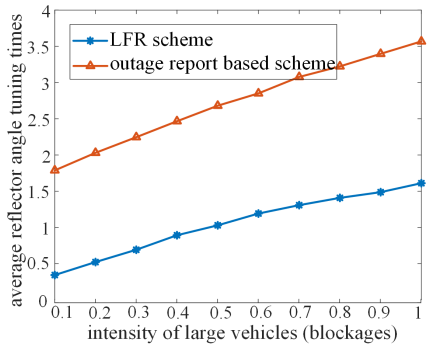

Fig. 7. Comparison of average reflector angle tuning units.

where the recovered data amount is more than three times under maximum blockages compared to that under minimum blockages, i.e., the comparison between the intensity of large vehicles under 1 and 0.1. In addition, since LFR scheme autonomously pre-tunes the angle of a reflector by outage predictions, we evaluate its performance by comparing with the outage-report based scheme, where the reflector angle is tuned based on the reported outages from blocked vehicles. We observe that LFR scheme outperforms the report based scheme by at most 10% more recovered data. To clarify this difference, we compare their reflector angle tuning time in each scheduling slot as a function of the intensity of large vehicles in Fig. 7. We observe that the reflector angle tuning time of the report based scheme is around 5 times as that of LFR scheme, when there are a few blockages. This is because the angle tuning in report based scheme sacrifices the time of data transmission, while LFR scheme can pre-tune the angle of a reflector and thus the tuning will sacrifice the time of data transmission time only when two different vehicles are blocked in consecutive time slots. This advantage of LFR scheme is diminished but still significant with more blockages.

V. Conclusion

In this paper, we have investigated how to implement tunable reflectors, the highly-reflective cheap metallic plates with adjustable angles, to augment mmWave V2I transmission environments for better communications, while not damaging the aesthetic nature of the environment. Specifically, we have first designed the system architecture, followed by the operational procedures. Considering the fundamental challenge of our design, i.e., tuning the angle of a reflector plate to adapt to the highly dynamic vehicular environments, we have proposed LFR algorithm based on DRL. Aiming at maximizing the system throughput, LFR autonomously learns from the observable traffic pattern to select the desirable reflector angle for the probably blocked vehicles in near future, in which way the LOS reflective links can be established in advance. Simulation results have demonstrated the effectiveness and the significant performance gain of our environment augmentation proposal.

Acknowledgment

This work was partially supported by US National Science Foundation under grants CNS-1717736 and IIS-1722791.

References

[1] J. Choi, V. Va, N. Gonzalez-Prelcic, R. Daniels, C. R. Bhat, and R. W. Heath, "Millimeter-wave vehicular communication to support massive automotive sensing," *IEEE Communications Magazine*, vol. 54, no. 12, pp. 160–167, 2016.

[2] Y. Wang, K. Venugopal, A. F. Molisch, and R. W. Heath, "Mmwave vehicle-to-infrastructure communication: Analysis of urban microcellular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 8, pp. 7086–7100, 2018.

[3] T. S. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5g cellular: It will work!" *IEEE access*, vol. 1, pp. 335–349, 2013.

[4] Q. Xue, X. Fang, and C.-X. Wang, "Beamspace su-mimo for future millimeter wave wireless communications," *IEEE Journal on Selected Areas in Communications*, vol. 35, no. 7, pp. 1564–1575, 2017.

[5] X. Zhou, Z. Zhang, Y. Zhu, Y. Li, S. Kumar, A. Vahdat, B. Y. Zhao, and H. Zheng, "Mirror mirror on the ceiling: Flexible wireless links for data centers," *ACM SIGCOMM Computer Communication Review*, vol. 42, no. 4, pp. 443–454, 2012.

[6] Z. Genc, U. H. Rizvi, E. Onur, and I. Niemegeers, "Robust 60 ghz indoor connectivity: Is it possible with reflections?" in *2010 IEEE 71st vehicular technology conference*. IEEE, 2010, pp. 1–5.

[7] Y. Niu, Y. Li, D. Jin, L. Su, and A. V. Vasilakos, "A survey of millimeter wave communications (mmwave) for 5g: opportunities and challenges," *Wireless Networks*, vol. 21, no. 8, pp. 2657–2676, 2015.

[8] L. Zhang, L. Yan, B. Lin, H. Ding, Y. Fang, and X. Fang, "Augmenting transmission environments for better communications: tunable reflector assisted mmwave wlans," http://www.fang.ece.ufl.edu/drafts/augmenting4journal.pdf, to be submitted for journal submission.

[9] R. W. Heath, "Millimeter wave communication: From origins to disruptive applications," http://users.ece.utexas.edu/~rheath/presentations/2017/MmWaveOriginsDisruptiveApplications_Lytle_Washington_2017.pdf, 2017.

[10] B.-G. Choi, W.-H. Jeong, and K.-S. Kim, "Characteristics analysis of reflection and transmission according to building materials in the millimeter wave band," *power (dBm)*, vol. 13, no. 50.62, pp. 64–48, 2015.

[11] N. C. Luong, D. T. Hoang, S. Gong, D. Niyato, P. Wang, Y.-C. Liang, and D. I. Kim, "Applications of deep reinforcement learning in communications and networking: A survey," *arXiv preprint arXiv:1810.07862*, 2018.

[12] H. Ye, Y. G. Li, and B.-H. F. Juang, "Deep reinforcement learning for resource allocation in v2v communications," *IEEE Transactions on Vehicular Technology*, 2019.

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

[14] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[15] P. A. Lopez, M. Behrisch, L. Bieker-Walz, J. Erdmann, Y.-P. Flötteröd, R. Hilbrich, L. Lücken, J. Rummel, P. Wagner, and E. WieBner, "Microscopic traffic simulation using sumo," in *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2018, pp. 2575–2582.