Frequency Prism-based Beamforming and Resource Allocation in LEO Satellite Networks with Delay Adjustable IRS

Shuta Sekimori[†1], Hiroaki Hashida[†‡2], Yuichi Kawamoto[†3], Nei Kato[†4],
Kohei Yoshida[§5], and Masayuki Ariyoshi[†§6]
[†]Graduate School of Information Sciences, Tohoku University, Sendai, Japan
[‡]Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai, Japan
[§]Advanced Network Research Laboratories, NEC Corporation, Kawasaki, Japan
E-mails: shuta.sekimori.s4@dc.tohoku.ac.jp[1], {hiroaki.hashida.d6[2], yuichi.kawamoto.d4[3], nei.kato.d3[4]}@tohoku.ac.jp,
{kohe-yoshida[5], m.ariyoshi[6]}@nec.com.

***Abstract*—This paper proposes an optimization strategy for multi-beam formation and time allocation control to meet the growing demand for low-Earth orbit (LEO) satellite communications and reduce communication latency. Conventional multi-beam forming, which sacrifices transmission power, has limitations in terms of further expansion. Therefore, we focused on frequency prism-based multi-beam formation with an antenna configuration that introduces a delay-adjustable intelligent reflecting surface (DA-IRS) with delay elements to address this limitation. These controls require time allocation control, which considers both current and future communication demands. Therefore, we propose a Q-learning-based control optimization method. The effectiveness of the proposed control optimization method is evaluated using simulations. Reducing the latency in LEO satellite communications will contribute to meeting expanding future demand.**

***Index Terms*—Q-learning, delay elements, frequency prisms, and intelligent reflecting surfaces**

I. Introduction

For sixth generation (6G), non-terrestrial network (NTN) systems are expected to ensure multimedia communications availability and scalability, providing wide-area coverage [1]. To achieve this, communication using low-Earth orbit (LEO) satellites, which have wide-area coverage, high speed, and low latency, has been considered [2]. LEO satellite communication requires antennas with high directional gain due to the high path loss. Traditionally, phased-array antennas have been employed [3], [4]. However, 6G networks are projected to employ higher frequency bands than those currently used by satellites. Achieving high transmission power at these frequencies with phased-array antennas results in significant antenna losses. Reflect arrays are used to address the limitations of antenna loss [5]. Antenna systems that reflect radio waves from a transmitter using a reflector array achieve high gain with low antenna loss. However, the phase shift of the reflect array is fixed, so the beam direction cannot be changed arbitrarily, and it cannot flexibly respond to dynamically changing channel conditions and communication requirements.

To address these limitations, we consider a LEO satellite equipped with an antenna in which an intelligent reflecting surface (IRS) reflects radio waves emitted from multiple transmitters [6]. An IRS controls the reflection characteristics of electromagnetic waves by using multiple metamaterial elements [7]. Each element adjusts the phase of the incoming waves, enabling beamforming in any direction and achieving high gain. The IRS passively reflects waves with low antenna loss, even at high frequencies, and can adapt to LEO satellite movement.

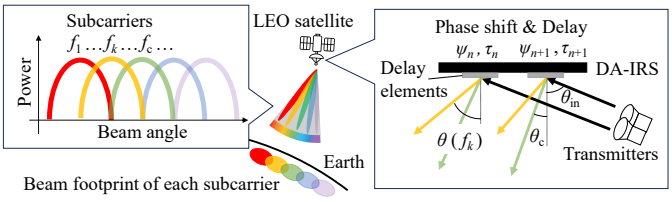

Fig. 1: Beamforming with frequency prism in LEO satellite communication system.

Additionally, LEO satellites must cover a wide area with low latency using beams with a limited spot range. If the satellite has a limited number of beams, covering a wide area through time division increases latency. Therefore, for low-latency multimedia communications, multi-beam forming has been considered to simultaneously cover a wide area [8], [9]. One method to form multiple beams with an IRS is to increase the number of transmitters that direct radio waves to the IRS. However, the challenge in increasing the number of transmitters is the corresponding increase in the transmission power. To address this issue, a multi-beam forming technology called a frequency prism has been considered [10], [11]. This technology introduces a delay-adjustable IRS (DA-IRS) equipped with delay elements [12] that can control the time until the radio waves incident on each element of the IRS are reflected. This disperses the beam direction of the multiple subcarriers multiplexed by orthogonal frequency-division multiplexing (OFDM), as shown in Fig. 1. This approach allows the formation of multiple beams without increasing the

979-8-3315-2096-0/25/$31.00 ©2025 IEEE

(a) Control to maximize communication volume for each time slot.

(b) Control to reduce long-term communication latency.

Fig. 2: Relationship between long-term communication latency and control.

transmission power. In [11], the control of the beam direction, the dispersion degree of the beam direction of the subcarriers, and the allocation of each subcarrier to different regions were managed for each control interval. However, as shown in Fig. 2 (a), controlling to maximize the communication volume at each time slot may increase the communication latency.

Therefore, in this paper, we propose a multi-beam control and resource allocation optimization method based on Q-learning in a frequency prism with a DA-IRS with the aim of reducing communication latency. The Q-learning-based method can make control decisions that reduce the long-term communication latency by learning the relationship between the distribution of communication demands, satellite control, and long-term communication latency, as shown in Fig. 2 (b). The effectiveness of this approach was evaluated for several communication-demand distributions.

The remainder of this paper is organized as follows. Section II describes the system model and provides an overview of the frequency prism, mechanisms by which the IRS realizes the frequency prism, and control methods employed. Section III presents an overview of Q-learning, the reasons for applying it, and the design methodology for its application. Section IV evaluates the performance of the Q-learning-based control decision-making method using simulations. Finally, Section V concludes the study and discusses future research.

II. System Model

In this section, we first explain the assumed system for LEO satellite communication. Next, we provide an overview of the principles behind the frequency prisms.

A. LEO Satellite Communication System

LEO satellite communication involves multiple satellites orbiting the Earth to meet communication demands. Each LEO satellite communicates with ground users within its coverage area. Each coverage area is divided into N_{grid} equally spaced grids, and there are communication demands $t_{x_i}^{\text{req}}$ of users in the i-th grid x_i. LEO satellites switch beam patterns every ΔT to satisfy communication demands across the entire grid. If all communication demands are completed by switching the beam control n_{switch} times in a certain communication-demand distribution, the latency time T_{latency} is defined as follows:

$$T_{\text{latency}} = n_{\text{switch}}\,\Delta T. \tag{1}$$

Each LEO satellite is equipped with one DA-IRS, which forms multiple beams using N_{antenna} antennas. Each beam is multiplexed with $N_{\text{subcarrier}}$ subcarriers using OFDM. The frequency prism disperses the beam directions of the $N_{\text{subcarrier}}$ subcarriers based on the beam direction of the center frequency, which is the middle frequency among all subcarriers, for each of the N_{antenna} beams, forming $N_{\text{subcarrier}} \times N_{\text{antenna}}$ multi-beam. In orthogonal frequency-division multiple access (OFDMA) communications using a frequency prism, different subcarriers are assigned to various grids and communicate simultaneously by loading arbitrary information on each subcarrier. The frequency prism controls the beam reflection direction of each subcarrier by controlling the delay of the $N_{\text{DA-IRS}}$ elements in the DA-IRS.

B. Frequency Prism by Satellite Mounted DA-IRS

A frequency prism is realized by sequentially reflecting radio waves incident on the DA-IRS from the elements at the end of the DA-IRS at different times. For simplicity, we assume that waves consisting of $N_{\text{subcarrier}}$ multiplexed subcarriers arrive at the DA-IRS from a single transmitter and consider a one-dimensional beam dispersion in the elevation direction θ. To disperse the beam for each subcarrier in the elevation direction using the DA-IRS frequency prism, the delay before the incident wave is reflected is configured for each row of elements. Here, the delay time τ_n of the n-th element row is set as follows:

$$\tau_n = (n-1)MT_{\text{c}}. \tag{2}$$

T_{c} is the period of the center frequency and M is the delay amount, which can be any positive integer. From (2), the phase-shift difference $\Delta\psi_\tau(f_k)$ between the elements for each frequency f_k is calculated as follows:

$$\Delta\psi_\tau(f_k) = 2\pi M\frac{(f_k - f_{\text{c}})}{f_{\text{c}}}, \tag{3}$$

where f_{c} denotes the center frequency. As indicated in (3), the delay time causes no phase difference between the elements at the center frequency. However, phase differences appear between the elements at frequencies other than the center frequency. This results in the beam direction varying for each subcarrier around the center frequency.

Additionally, the delay elements of the DA-IRS not only control the delay time but also set a phase shift $\psi_n(\theta_{\text{c}})$ for

the n-th element row across all frequencies, ensuring that the center frequency is reflected in the desired direction θ_c. This can be expressed as follows:

$$\psi_n(\theta_c) = \frac{2\pi f_c}{c} d(n-1)(\sin\theta_c - \sin\theta_{in}), \quad (4)$$

where c is the speed of light, d is the spacing between elements of the DA-IRS, and θ_{in} is the angle of incidence of the DA-IRS. From (3) and (4), the phase shift of the n-th element row of the DA-IRS can be expressed as follows:

$$\psi_n^{\text{DA-IRS}} = \psi_n(\theta_c) + (n-1)\Delta\psi_\tau(f_k). \quad (5)$$

Next, we model the beam pattern that spreads out in response to each subcarrier of the incident wave. The beam pattern of the frequency prism results from the superposition of signals reflected by all the DA-IRS elements and can be represented as an array factor (AF):

$$g_{\text{AF}}(f_k,\theta) = e^{i\frac{\sigma}{2}(N_{\text{DA-IRS}}-1)} \frac{\sin\frac{\sigma}{2} N_{\text{DA-IRS}}}{\sin\frac{\sigma}{2}}, \quad (6)$$

where $\sigma = -\frac{2\pi M}{f_c}(f_k - f_c) - \frac{2\pi f_c d}{c}(\sin\theta_c - \sin\theta_{in}) + \frac{2\pi f_k d \sin\theta}{c}$. The variable $g_{\text{AF}}(f_k,\theta)$ represents the amplitude in the direction of the reflection θ when a radio wave with frequency f_k and an amplitude of 1 enters the DA-IRS and is reflected under the condition that the phase shift amount of the DA-IRS satisfies (5). For simplicity, the reflection coefficient of the DA-IRS is assumed to be 1 in this study. The beam direction $\hat{\theta}(f_k)$ of frequency f_k is θ such that $\sigma = 0$, as shown below:

$$\hat{\theta}(f_k) = \sin^{-1}\left\{ \frac{Mc(f_k - f_c)}{f_k f_c d} + \frac{f_c}{f_k}(\sin\theta_c - \sin\theta_{in}) \right\}. \quad (7)$$

The direction of the beam varies for each subcarrier, enabling frequency-based multi-beam forming. Furthermore, the degree of dispersion of each subcarrier based on the center frequency can be controlled by adjusting the delay amount M.

III. Q-Learning-Based Beamforming and Resource Allocation

In this paper, we propose a Q-learning-based optimization method for reducing long-term communication latency in multi-beam forming and resource allocation in the frequency prism using DA-IRS. In this section, we provide an overview of the proposed method and the reasons for applying Q-learning. Next, we describe the Q-learning design and the method used to make control decisions.

A. Overview of Proposed Method

Q-learning is a method that empirically learns the optimal actions by considering future changes in a situation and constructing a Q-table [13]. In this study, we propose a control method that can reduce long-term communication latency using a Q-table pre-learned from the distribution of ground communication demands at a given time t. To reduce communication latency, it is necessary to increase the amount of communication that can be completed in each control. However, if control is performed to complete most of the communication in each time slot, the distribution of demands in the next time slot may become sparse, which may reduce the amount of communication that can be completed in the subsequent control. This is because, as shown in Fig. 3, a sparse distribution of communication demands reduces the number of available subcarriers when simultaneously covering multiple distant demands since the frequency prism cannot generate branching beams. In this study, we apply Q-learning to learn the relationship between the density of the distribution of communication demands, the amount of communication that can be completed with one control, and long-term communication latency. This makes it possible to determine a control that reduces long-term communication latency for each time slot.

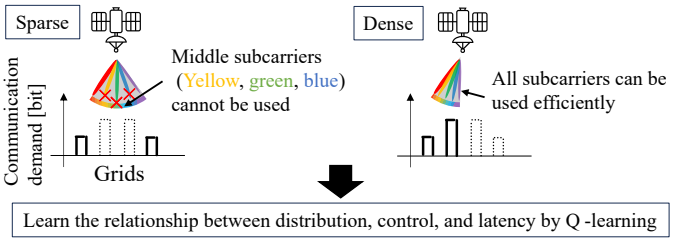

Fig. 3: Relationship between density of communication-demand distribution and available subcarriers.

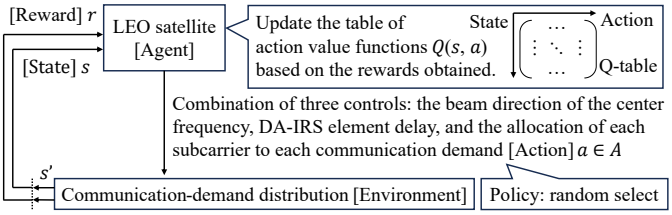

Fig. 4: Q-learning-based beamforming and resource allocation scheme.

B. Q-Learning Scheme

First, we explain the application of Q-learning. An overview of Q-learning is presented in Fig. 4. In Q-learning, an agent and an environment are set up to solve a problem through action. In LEO satellite communication, multiple LEO satellites orbit Earth. However, as all LEO satellites perform Q-learning and make control decisions in the same manner, an agent is defined as a single LEO satellite. Each LEO satellite obtains ground communication-demand distribution at time t, which is defined as the Q-learning environment. The agent discretizes the communication demands in the environment, which are continuous values, and defines them as states. Each LEO satellite then determines three controls as the Q-learning actions in the control cycle: the beam direction of the center frequency, the DA-IRS element delay, and the allocation of each subcarrier to each communication demand. In the learning phase, when an action is selected from a certain state, a reward is provided according to the effect of that action on the communication latency time, and the Q-table is

updated. This Q-table contains Q-values, which are the values of each action in each state. In the execution phase, the agent already has the Q-table updated in the learning phase to reduce the long-term communication latency and does not update it. Communication latency is minimized by selecting an action that maximizes the Q-value for each control interval until all communication demands are fulfilled.

Next, we formulate the state and actions in Q-learning and explain the transition to the next state when a certain action is performed in a certain state. The Q-learning environment is a communication-demand distribution with the communication demand $t_{x_i}^{\text{req}}$ in each communication-demand grid x_i. State s at time t is as follows:

$$s = [s_{x_0}, ..., s_{x_i}, ..., s_{x_{N\,\text{grid}}}], \tag{8}$$

where s_{x_i} is the state of the communication demand in each area, defined as follows:

$$s_{x_i} = \begin{cases} 0, & \text{if } t_{x_i}^{\text{req}}(t) = 0, \\ s_{\max}, & \text{if } s_{\max} - 0.5 \leq t_{x_i}^{\text{req}}(t), \\ \max\left(1, \text{round}(t_{x_i}^{\text{req}}(t))\right), & \text{else}, \end{cases} \tag{9}$$

where $s_{\max}$ denotes the maximum state value in each grid. Thus, the size $|S|$ of the state space is expressed as follows:

$$|S| = {s_{\max}}^{N_{\text{grid}}}. \tag{10}$$

The action at time t is a combination of three controls: the beam direction θ_{c} of the center frequency, DA-IRS element delay M, and allocation H of each subcarrier to each communication-demand grid, as follows:

$$a = [\theta_{\text{c}}(t), M(t), H(t)]. \tag{11}$$

The range that each control element can assume is preset to a finite number, and the set of these combinations is defined as an action pattern A, which is $a \in A$. The size $|A|$ of this action pattern A is the size of the action space for Q-learning, and is expressed as follows:

$$|A| = N_{\theta_{\text{c}}} \times N_M \times N_H, \tag{12}$$

where $N_{\theta_{\text{c}}}$, N_M, and N_H are the number of patterns in the center frequency beam direction, the DA-IRS element delay, and the allocation of each subcarrier to each communication-demand grid, respectively, which is the number of overlapping combinations of the number of communication-demand grids N_{grid} and the number of subcarriers $N_{\text{subcarrier}}$. From (10) and (12), the Q-table size is $|A| \times |S|$.

Finally, we explain how to update the Q-table to reduce the long-term communication latency in the learning phase. When an agent is given an environment with state s, it randomly selects action a from a set of predefined action patterns A to determine control. When subcarrier k is allocated to communication-demand grids x_i by a certain control at time t by $H(t)$, the communication demand $t_{x_i}^{\text{req}}(t + \Delta T)$ in area x_i after control switching is expressed as follows:

$$t_{x_i}^{\text{req}}(t + \Delta T) = \max\left(0,\, t_{x_i}^{\text{req}}(t) - \Delta T \cdot t_{k,x_i}^{\text{ach}}\right). \tag{13}$$

TABLE I: Settings of three distributions

	Communication demand for each grid in each case [Gbit]				
	-165 km	-82.5 km	0 km	82.5 km	165 km
Case 1	3	4	6	3	3
Case 2	3	4	9	4	4
Case 3	9	4	9	4	7

t_{k,x_i}^{ach} is the throughput that subcarrier k can achieve at position x_i and is expressed as follows:

$$t_{k,x_i}^{\text{ach}} = B \log_2(1 + \gamma_{k,x_i}(\theta_{\text{c}}, M)), \tag{14}$$

where B is the bandwidth of each subcarrier, and $\gamma_{k,x_i}(\theta_{\text{c}}, M)$ is the signal-to-noise ratio (SNR) that subcarrier k achieves in grid x_i when the central frequency beam direction is θ_{c} and the DA-IRS element delay is M. Thus, when action a is taken from time t to time $t+\Delta T$, the communication demand $t_{x_i}^{\text{req}}(t)$ of each grid changes, causing state s to transition to state s'. At this time, a reward r is assigned: 1 if all communication demands are completed, -1 if no communication occurs with action a; otherwise, 0. Thus, the action value function $Q(s, a)$ can be updated using the following formula:

$$Q(s, a) \mathrel{+}= \alpha \left[r + \gamma \max_{a \in A} Q(s', a) - Q(s, a) \right], \tag{15}$$

where α is the learning rate and γ is the discount rate. By setting γ to $0 < \gamma < 1$, Q-learning minimizes long-term delay time. This update is repeated until all communication demands are completed, defining the sequence of actions and updates as one episode. By conducting multiple episodes of learning, the Q-table is continuously updated to reduce long-term communication latency. It is also possible to improve the learning efficiency by preparing a simulation environment for each episode and performing preliminary learning.

IV. Evaluation

The purpose of the simulation is to evaluate the control decision method using Q-learning with respect to long-term communication latency. In this simulation, we assume that in OFDMA communications, the three subcarriers emitted from a single transmitter are directed toward the DA-IRS and reflected toward the ground. On the ground, we only consider the left-right direction with the grid divided into five parts and evaluate three different cases, as shown in Table I. Directly below the satellite is considered to be 0 km, and the grids are defined as positive and negative on both sides. In each case, the control is switched at every control interval ΔT until the communication demand $t_{x_i}^{\text{req}}$ in the five communication-demand grids becomes zero, and the communication latency time is evaluated. Without loss of generality and for simplicity, the change in the relative position with respect to the ground owing to the movement of the LEO satellite is not considered in this simulation. We compare the proposed method with a method that repeats the control that completes most communication demands in each time slot (Comp 1) and a method that randomly selects a control from a control pattern (Comp 2). In Comp 2, the latency averaged over 100 iterations

TABLE II: List of parameter settings

Parameter	Definition
Transmit power	20 dB
Noise power	-100 dB
Center frequency	10.5 GHz
Bandwidth	0.25 GHz
Number of DA-IRS elements	1024
Altitude of a satellite	800 km
Control cycle	1 sec
Learning rate	0.5
Discount rate	0.9
Number of episodes for each case	1,000,000

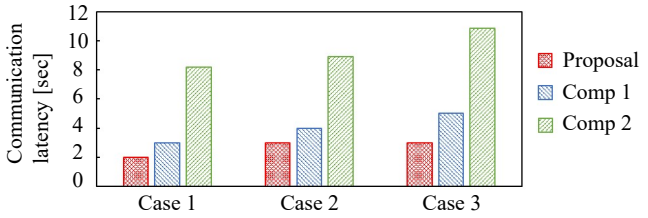

Fig. 5: Communication latency of three methods in three cases.

is evaluated for each case. In the proposed method, the control that maximizes the Q-value is selected using a Q-table trained for N_{episode} in each case. As states 0 to 9 are prepared for each of the five grids, the size of the state space is 10^5. There are five center frequency directions corresponding to the five communication-demand grid directions. The delay is set to 10 values ranging from 0 to 9. The allocation of each subcarrier to each communication demand has 125 possible combinations with overlapping combinations because there are five communication-demand grids and three subcarriers. The parameters are listed in Table II [14].

The results are presented in Fig. 5. The proposed method minimizes the long-term communication latency in all three cases. This is because the proposed method switches the control while maintaining the dense distribution of communication demand. Conversely, Comp 1 controls the communication demands completed in the first time slot to be higher; therefore, the communication-demand distribution in subsequent time slots becomes sparse and less communication demand can be completed. For example, in Cases 1 to 3, Comp 1 communicates with the 0 km grid, where the communication demand is concentrated in the first time slot, while the proposed method communicates with grids other than the 0 km grid. In the next time slot, in Comp 1, the communication demand is distributed outside the 0 km grid and becomes sparse, whereas in the proposed method, the communication demand remains dense around the 0 km grid. Furthermore, comparing Cases 2 and 3, the latency increases in Comp 1 but not in the proposed method. This is because in Case 3, the communication demand is distributed more in the $\pm$165 km grid than in Case 2; therefore, the communication-demand distribution becomes sparser after Comp 1 communicates with the 0 km grid in the first time slot. These results show that the proposed method minimizes communication latency and meets communication demands faster.

V. Conclusion And Future Directions

In this study, we propose a control decision method using Q-learning to achieve low-latency communication in multiple beamforming in a frequency prism using a DA-IRS integrated into a LEO communication satellite. We conducted simulations to evaluate the proposed method and confirmed that it reduced communication latency in a specific environment. These results indicate the effectiveness of our approach in meeting the future expanding communication demands of LEO satellite communication systems. However, if Q-learning is performed over a wider area, such as the coverage of neighboring satellites, the number of states becomes enormous, making learning difficult. Therefore, future studies will consider applying deep Q-networks that can handle continuous-state values in the environment.

References

[1] M. Giordani and M. Zorzi, "Non-terrestrial networks in the 6G era: Challenges and opportunities," *IEEE Netw.*, vol. 35, no. 2, pp. 244–251, Mar. 2021.

[2] A. Kapovits, M.-I. Corici, I.-D. Gheorghe-Pop, A. Gavras, F. Burkhardt, T. Schlichter, and S. Covaci, "Satellite communications integration with terrestrial networks," *China Commun.*, vol. 15, no. 8, pp. 22–38, Aug. 2018.

[3] A. Ivanov, R. Bychkov, and E. Tcatcorin, "Spatial resource management in LEO satellite," *IEEE Trans. Veh. Technol.*, vol. 69, no. 12, pp. 15 623–15 632, Dec. 2020.

[4] X. Fang, W. Feng, T. Wei, Y. Chen, N. Ge, and C.-X. Wang, "5G embraces satellites for 6G ubiquitous IoT: Basic models for integrated satellite terrestrial networks," *IEEE Internet of Things J.*, vol. 8, no. 18, pp. 14 399–14 417, Sep. 2021.

[5] S. Anguix, A. Araghi, M. Khalily, and R. Tafazolli, "Reflectarray antenna design for LEO satellite communications in Ka-band," in *2021 15th Eur. Conf. on Antennas and Propag. (EuCAP)*, Mar. 2021, pp. 1–5.

[6] K. Tekbıyık, G. K. Kurt, and H. Yanikomeroglu, "Energy-efficient RIS-assisted satellites for IoT networks," *IEEE Internet of Things J.*, vol. 9, no. 16, pp. 14 891–14 899, Aug. 2022.

[7] H. Hashida, Y. Kawamoto, and N. Kato, "Intelligent reflecting surface placement optimization in air-ground communication networks toward 6G," *IEEE Wireless Commun.*, vol. 27, no. 6, pp. 146–151, Dec. 2020.

[8] R. Wu and F. Yang, "Study on key technique and research hotspot of low orbit satellite communication," in *2021 Int. Conf. Wireless Commun. Smart Grid (ICWCSG)*, Aug. 2021, pp. 89–93.

[9] J. Choi and V. Chan, "Satellite multibeam allocation and congestion control with delay constraints," in *2004 IEEE Int. Conf. Commun. (IEEE Cat. No.04CH37577)*, vol. 6, Jun. 2004, pp. 3309–3315.

[10] B. Zhai, Y. Zhu, A. Tang, and X. Wang, "Thzprism: Frequency-based beam spreading for terahertz communication systems," *IEEE Wireless Commun. Lett.*, vol. 9, no. 6, pp. 897–900, Jun. 2020.

[11] S. Sekimori, Y. Kawamoto, N. Kato, S. Watanabe, J. Funada, and M. Ariyoshi, "Frequency prism in delay adjustable intelligent reflecting surfaces for long distance communications in LEO satellite networks," *2024 IEEE Int. Conf. Commun. (ICC)*, in press.

[12] T. Nakanishi and M. Kitano, "Storage and retrieval of electromagnetic waves using electromagnetically induced transparency in a nonlinear metamaterial," *Appl. Phys. Lett.*, vol. 112, no. 20, p. 201905, May. 2018. [Online]. Available: https://doi.org/10.1063/1.5035442

[13] Y. Qian, J. Wu, R. Wang, F. Zhu, and W. Zhang, "Survey on reinforcement learning applications in communication networks," *J. Commun. Inf. Networks*, vol. 4, no. 2, pp. 30–39, Jun. 2019.

[14] W. Tang, M. Z. Chen, X. Chen, J. Y. Dai, Y. Han, M. Di Renzo, Y. Zeng, S. Jin, Q. Cheng, and T. J. Cui, "Wireless communications with reconfigurable intelligent surface: Path loss modeling and experimental measurement," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 421–439, Jan. 2021.