Computer Vision-Aided Reconfigurable Intelligent Surface-Based Beam Tracking: Prototyping and Experimental Results

Ming Ouyang, Graduate Student Member, IEEE, Feifei Gao, Fellow, IEEE, Yucong Wang, Graduate Student Member, IEEE, Shun Zhang, Senior Member, IEEE, Puchu Li, and Jian Ren, Member, IEEE

Abstract—In this paper, we propose a novel computer vision-based approach to aid reconfigurable intelligent surface (RIS) for dynamic beam tracking and implement the corresponding prototype verification system. A camera is attached at the RIS to obtain the visual information about the surroundings, with which RIS identifies the incident beam direction and the desired reflected beam direction, and then adjusts the reflection coefficients according to the pre-designed codebooks. We build a 20-by-20 RIS running at 5.4 GHz and develop a high-speed control board to ensure the real-time refresh of the reflection coefficients. Meanwhile, we implement an independent peer-to-peer communications system to simulate the physical link between the base station and the user equipment. The vision-aided RIS prototype system is tested in two mobile scenarios: RIS works in near-field conditions as a passive array antenna of the base station; RIS works in far-field conditions to assist the communication between the base station and the user equipment. The experimental results show that RIS can quickly adjust the reflection coefficients for dynamic beam tracking with the help of visual information.

Index Terms—Reconfigurable intelligent surface (RIS), prototype system, computer vision, beam tracking.

I. INTRODUCTION

RECONFIGURABLE intelligent surface (RIS) assisted communications has emerged as one of the main technologies for the next-generation mobile communication systems and can yield significant improvements at a cheap cost [1], [2], [3], [4], [5]. A RIS is a uniform array of elements that can modulate the incident wave’s amplitude and phase. The primary principle of RIS is to leverage the array elements’ modulation abilities to generate certain reflected beams, thereby enhancing the signal power in the specific directions. Different RIS codewords determine different directions of the reflected beam, and the user’s position is not deterministic, thus the key challenge is to design the appropriate reflection coefficients for dynamic beam tracking. Currently, there are three main ways to compute the reflection coefficients: channel state information (CSI)-based schemes, beam-sweeping-based schemes, and end-to-end deep learning-based schemes.

The CSI-based schemes can be further divided into deterministic CSI-based [6], [7], [8], [9], [10], [11] and statistical CSI-based schemes [12], [13], [14]. In [6], the authors selected the optimal reflection coefficients from a pre-designed codebook after estimating the overall equivalent channel such that the spectral efficiency can be maximized. In [7], the authors designed a low overhead majorization-minimization-based method to optimize the reflection coefficients. In [8], the authors divided the channel into multiple subchannels, each corresponding to a RIS array element, and calculated the reflection coefficients based on the CSI of each subchannel. In [9], [10], and [11], the authors introduced compressed sensing into the channel estimation process and calculated the reflection coefficients based on the recovered sparse CSI. Typically, the statistical CSI is easier to obtain than the deterministic CSI. In [12], [13], and [14], the authors optimized the reflection coefficients based on the statistical CSI and analyzed the performance of the optimization algorithms. Although it is very effective to design reflection coefficients based on the CSI, the huge number of elements on RIS leads to a large training overhead.

Compared with the CSI-based schemes, the beam-sweeping-based schemes does not require complex channel estimation and therefore are highly advantageous for RIS with a large number of array elements [15], [16], [17]. In [15], the authors utilized a manifold-based algorithm to calculate the optimal reflection coefficients for specific directions and then took beam-sweeping scheme to obtain the correct direction. In [16] and [17], the authors validated the beam-sweeping scheme in the real-world environment. In order to speed up the sweep, the authors of [18] introduced a greedy algorithm to
search for the optimal reflection coefficients. However, these beam-sweeping schemes not only consumes a large amount of sweep time, but also requires extra feedback links, which enlarges the communications delay.

In end-to-end deep learning-based schemes, [19] and [20] leveraged neural networks to output the optimal reflection coefficients directly. However, the neural network in [19] requires the user’s location information as input, and the neural network in [20] requires the power distribution near the user as input. The acquisition of the location information or the power distribution likewise results in significant additional training overhead.

Recently, the out-of-band information is introduced into the communication systems to reduce the beam training overhead and eliminate the feedback links, e.g., [21], [22], [23], [24], [25], and [26]. In [21], the sub-6 GHz channel covariance was applied to assist the estimation of the mmWave channel covariance. In [22], MIMO radar was utilized to aid mmWave base station in estimating the time-varying channel. In [23], [24], [25], and [26], visual information was leveraged to assist beam tracking, blockage prediction or proactive handoff. For RIS, choosing the visual information may be a much more suitable option due to the following three reasons: (1) RIS is mainly used to assist communications through reflected line of sight (LOS) path, especially for mmWave band and Terahertz band, whereas the camera is being precisely effective for quickly locating users within LOS; (2) One of the advantages of RIS is its low cost, and therefore it is appropriate to replace the complex feedback links with a cheap camera; (3) The acquisition of visual information does not occupy other frequency bands, which saves spectrum resources and beam training overhead. However, in contrast to conventional vision-aided communication systems, vision-aided RIS, as a third-party auxiliary device, is loaded with codewords that are determined by both the desired emitted beam direction and the incident beam direction. Thus the number of codewords is quadratic to that of conventional communication systems, which makes the storage of codewords difficult. Moreover, although the visual algorithms can acquire the surrounding information in real time, the huge number of array elements of RIS makes the real-time refreshing of codewords difficult, which limits the advantages of visual assistance. Motivated by these challenges, we propose an effective computer vision-based scheme to address codeword storage problem and implement a high-speed RIS control board to enable the real-time refreshing of the RIS codewords. The main contributions of this paper can be summarized as follows:

- For the first time, an effective computer vision-based approach is proposed to assist RIS in realizing beam tracking. A camera is attached at the RIS to obtain the visual information about the surroundings, with which RIS identifies the location of transmitter and the direction of receiver. Then RIS determines the codebook based on the transmitter’s location and the codeword based on the receiver’s direction, which addresses codeword storage challenge.
- A high-speed RIS control board is developed to enable the real-time refresh of the reflection coefficients.

We cascade two low-cost, high-I/O-density FPGA chips (Intel Cyclone IV EP4CE15F23C8N) in the RIS control board, which can run at up to 200MHz system clock frequency and control each PIN diode individually. Compared with the common shift register solution [16], the codeword refresh speed of the designed control board is several times faster. Moreover, the method of cascading is more easily reusable for RIS with a larger number of units.

- The proposed prototype system is tested in two classical scenarios: RIS works in near-field conditions as a passive array antenna of the base station; RIS works in far-field conditions to assist the communications between the base station and the user equipment. The experimental results show that RIS can quickly adjust the reflection coefficients for dynamic beam tracking with the help of visual information.

The remainder of the paper is composed of the following parts: Section II presents the system model and the design principle of the reflection coefficients. Section III provides the specific implementation details of vision-based beam tracking scheme. Section IV describes the vision-aided RIS prototype verification system, including the design of the RIS codebook, the control board, and the architecture of the peer-to-peer communications system. In Section V, we test the vision-aided RIS in two scenarios and analyze the benefits of visual information. In Section VI, we analyze the beam tracking performance as well as the limitations of the prototype system. Finally, in Section VII, we draw the conclusions.

II. SYSTEM MODEL

We assume that the base station (BS) and RIS serve only one user equipment (UE) in the communications system, as shown in Fig. 1. Moreover, the BS and UE are equipped with single antenna, while the RIS contains $M \times N$ units. The orthogonal frequency division multiplexed (OFDM) is adopted and the baseband contains $K$ subcarriers. Let $h_{tx}[k], h_{rx}[k] \in \mathbb{C}^{MN \times 1}$ represent the channel between the UE and RIS, and that between the BS and RIS at the $k$-th subcarrier, respectively. Denote $H_{RIS} \in \mathbb{C}^{MN \times MN}$ as the manipulation matrix at RIS. Note that $H_{RIS}$ is a diagonal matrix, whose diagonal elements can form the vector
\[ \mathbf{h}_{\text{ris}} = [A_{11}e^{j\alpha_{11}}, \cdots, A_{1N}e^{j\alpha_{1N}}, \cdots, A_{MN}e^{j\alpha_{MN}}]^T, \]

where \( A_{mn}, \alpha_{mn} \) represent the modulated phase and modulated amplitude of the \( mn \)-th unit in RIS. We assume the line of sight (LOS) path between BS and UE is blocked, and then the UE’s received signal \( r_k \) at the \( k \)-th subcarrier can be written as

\[ r_k = h_{rx}^T[k]\mathbf{H}_{\text{RIS}}h_{tx}[k]s_k + n_k, \]

where \( s_k, n_k \in \mathbb{C} \) denote the BS’s transmitting signal and the noise. The channel capacity \( C \) can be computed as

\[ C = \sum_{i=1}^{K} \log_2 \left( 1 + \frac{|h_{rx}^T[k]\mathbf{H}_{\text{RIS}}h_{tx}[k]s_k|^2P_k}{\sigma^2} \right), \]

where \( P_k, \sigma^2 \) represent the average power of the signal and the variance of the noise, respectively.

The manipulation matrix \( \mathbf{H}_{\text{RIS}} \) should be designed to maximize the channel capacity \( C \). According to (2), we can obtain the approximate optimal beamforming vector \( \mathbf{h}_{\text{ris}}^{\text{optm}} \) as

\[ \mathbf{h}_{\text{ris}}^{\text{optm}} = \arg \max_{A_{mn}, \alpha_{mn}} \prod_{i=1}^{K} |h_{rx}^T[k]\text{diag}(\mathbf{h}_{\text{ris}})h_{tx}[k]s_k|^2P_k. \]  

According to (3), the prerequisite for calculating \( \mathbf{h}_{\text{ris}}^{\text{optm}} \) is to obtain the CSI \( h_{rx}[k] \) and \( h_{tx}[k] \). However, the large number of array units on the RIS poses a great challenge for channel estimation.

In order to deal with the challenge of channel estimation, we transform (3) into maximizing the received signal power at the UE. We establish a coordinate system with the center point of the RIS board as the coordinate origin and use it as the phase reference point, as shown in Fig. 2. According to [27], the scattering field pattern function of RIS can be expressed as

\[ E(\theta, \varphi) = \sum_{m=1}^{M} \sum_{n=1}^{N} f_{mn}(\theta, \varphi)A_{mn}B_{mn}\exp(j(\alpha_{mn} + \beta_{mn})) \]

\[ \times \exp \left( j \left( \frac{2\pi d}{\lambda} \left( \frac{M + 1}{2} - m \right) \cos \theta \right) \right) \]

\[ \times \exp \left( j \left( \frac{2\pi d}{\lambda} \left( n - 1 + \frac{N}{2} \right) \sin \theta \sin \varphi \right) \right), \]

where \( \theta, \varphi \) are the pitch and azimuth angles relative to RIS, \( f_{mn}(\theta, \varphi) \) stands for the scattering field pattern of the \( mn \)-th unit in RIS, \( B_{mn}, \beta_{mn} \) stand for the amplitude and phase of the incident wave at the \( mn \)-th unit, while \( \lambda \) and \( d \) correspond to the wavelength of electromagnetic wave and the length of spacing between RIS units respectively. Specifically, the value of \( \beta_{mn} \) depends on the location of the transmitter relative to the RIS. The relationship between the antenna gain of RIS in each direction and the scattering pattern is given by [17]

\[ G_{\text{RIS}}(\theta, \varphi) \propto |E(\theta, \varphi)|^2. \]

Denote \( (\theta_{rx}, \varphi_{rx}) \) as the direction of the UE. Then we can obtain the relationship between the receiving power \( P_r \) at UE and the transmitting power \( P_t \) at BS as:

\[ P_r = \sum_{k=1}^{K} P_k G_t G_{\text{RIS}}(\theta_{rx}, \varphi_{rx})G_r \left( \frac{\lambda}{4\pi D} \right)^2, \]

\[ \text{Fig. 2. Coordinate system established with the RIS center as the origin,} \]

where \( G_t \) denotes the gain of transmitting antenna, \( G_r \) denotes the gain of receiving antenna, and \( D \) stands for the distance travelled by electromagnetic wave. In addition, \( (\lambda/4\pi D) \) is the free space path loss (FSPL) of electromagnetic wave. Typically, \( A_{mn} \) is a constant, and thus we only need to consider the phase modulation of the incident wave. Let us define \( h_{\text{ris}, \alpha}^{\text{optm}} \) as the vector consisting of the optimal modulated phase of RIS units, i.e., \( h_{\text{ris}, \alpha}^{\text{optm}} = [\alpha_{11}^{\text{optm}}, \cdots, \alpha_{1N}^{\text{optm}}, \cdots, \alpha_{MN}^{\text{optm}}] \). Then (3) can be rewritten as

\[ \mathbf{h}_{\text{ris}, \alpha}^{\text{optm}} = \arg \max_{\alpha_{mn}} P_r. \]

According to (5), (6) and (7), equation (3) eventually translates into designing a proper modulated phase \( \alpha_{mn} \) to maximize \( E(\theta_{rx}, \varphi_{rx}) \). Then \( \mathbf{h}_{\text{ris}, \alpha}^{\text{optm}} \) can be expressed as

\[ \mathbf{h}_{\text{ris}, \alpha}^{\text{optm}} = \arg \max_{\alpha_{mn}} E(\theta_{rx}, \varphi_{rx}). \]

According to (4) and (8), we can calculate the optimal modulation phase as

\[ \alpha_{mn}^{\text{optm}} = -2\pi \left( \left( \frac{M + 1}{2} - m \right) \cos \theta_{rx} \right) \]

\[ -2\pi \left( \left( n - 1 + \frac{N}{2} \right) \sin \theta_{rx} \sin \varphi_{rx} \right) - \beta_{mn}. \]

Traditionally, we can take the exhaustive search method or beam sweeping method to obtain \( (\theta_{rx}, \varphi_{rx}) \) and \( \beta_{mn} \). However, beam sweeping scheme would incur a huge training overhead, which greatly affects the quality of the communications. We next introduce a novel method of using visual information to assist RIS in obtaining \( (\theta_{rx}, \varphi_{rx}) \) and \( \beta_{mn} \), which can greatly save the beam training overhead.

### III. Vision-Based Beam Tracking Scheme

In the proposed vision-based beam tracking scheme, we use a binocular camera to obtain visual information of the target.
device (UE or BS) with advanced object detection algorithm as well as the stereo vision algorithm. The object detection algorithm can provide the device’s 2D coordinates in the image plane, while the stereo vision algorithm can calculate the device’s 3D coordinates with respect to RIS. The proposed beam tracking scheme can be divided into two stages. In the first stage, the matching RIS codebook can be determined based on the 3D coordinates of the BS with respect to RIS. In the second stage, the direction of the UE relative to RIS is calculated based on the real-time 3D coordinates of the BS. Then, based on the 3D coordinates of the BS, we can obtain the environment image from camera, the backbone network CSPDarknet [31] performs feature extraction on the image. The extracted multi-scale features are then fed into the feature pyramid network (FPN) and path aggregation network (PAN) for feature fusion and enhancement. Then three decoupled heads are used for regression and classification, and the output is decoded to obtain the prediction information \( \{ x_c, y_c, w, h \} \) of the target device in the image. Note that \( x_c, y_c \) represent the 2D coordinates of the center point of the prediction bounding box while \( w, h \) denote the width and height of the prediction bounding box.

However, the object detection algorithm can only provide the 2D coordinates of the target device in the image. We then adopt the binocular stereo vision algorithm to calculate the 3D coordinates in the real world, that includes the following two steps: Firstly, stereo matching algorithm is used to obtain the vision disparity between the two images \([32],[33]\). Then the depth information can be calculated based on the vision disparity. Taking the calculation of UE’s visual information as an example, the depth information computation model of the binocular camera is shown in Fig. 4, where \((x_L, y_L), (x_R, y_R)\) denote the coordinates of the left and the right image centroids, \((x_1, y_1), (x_2, y_2)\) denote the 2D coordinates of the UE in two images, \((x_{rx}, y_{rx}, z_{rx})\) denote the 3D coordinates of the UE with respect to RIS, and \(f, b\) denote the focal length and the spacing between the two cameras. Typically, the vision disparity \(d_v\) is defined as

\[
d_v = x_1 - x_L + x_R - x_2. \tag{10}
\]

Based on the binocular geometry \([34]\), we can derive the depth \(z_{rx}\) of the UE as

\[
z_{rx} = \frac{fb}{d_v}. \tag{11}
\]

Combining the depth \(z_{rx}\), the intrinsic matrix of the camera and the 2D coordinates \((x_c, y_c)\) output by YOLOX, we can calculate the 3D coordinates \((x_{rx}, y_{rx}, z_{rx})\) of the UE with respect to RIS. In the same way, we can also calculate the 3D coordinates \((x_{tx}, y_{tx}, z_{tx})\) of the BS. Denote \(\theta_{rx}, \varphi_{rx}\) as the the pitch angle and the azimuth angle of the UE with respect to RIS, respectively. Given \(\theta_{rx}, \varphi_{rx} \in [0, \pi]\) and \(\varphi_{rx} \in [-\frac{\pi}{2}, \frac{\pi}{2}]\), the direction \((\theta_{rx}, \varphi_{rx})\) of the UE can be calculated as:

\[
\theta_{rx} = \begin{cases} 
\arctan \left( \frac{\sqrt{x_{rx}^2 + z_{rx}^2}}{y_{rx}} \right) & y_{rx} > 0, \\
\pi + \arctan \left( \frac{\sqrt{x_{rx}^2 + z_{rx}^2}}{y_{rx}} \right) & y_{rx} < 0,
\end{cases} \tag{12}
\]

\[
\varphi_{rx} = \arctan \left( \frac{x_{rx}}{z_{rx}} \right). \tag{13}
\]

Based on the 3D coordinates of the BS, we can obtain the phase \(\beta_{mn}\) of the incident beam on each RIS unit, and thus determine the corresponding codebook. Then, based on the
direction \((\theta_{rx}, \varphi_{rx})\) of the UE, we can quickly adjust the optimal codeword of RIS to realize beam tracking.

IV. SYSTEM DESIGN

The architecture of the proposed prototype system is shown in Fig. 5, where Fig. 5(a) presents the theoretical block diagram and Fig. 5(b) presents the physical diagram. In this section, we clarify the system design in four aspects, including vision module, RIS and codebook, control board, and peer-to-peer communications system.

A. Vision Module Design

In the prototype system, both the UE and the BS are equipped with horn antennas. Therefore, the vision module calculates the specific location of the horn antenna as the location of the UE and the BS. In order for YOLOX to accurately detect the horn antenna, we create a dataset of 1000 photos containing the horn antenna and train YOLOX loaded with pre-trained weights. After the training process, YOLOX is able to successfully detect the location of the horn antenna in the photos returned from the binocular camera in real time.

The binocular camera adopted is the ZED II camera provided by Stereolabs, that takes deep learning-based stereo matching algorithm to calculate the vision disparity and then calculates the 3D coordinates of the pixels. In addition, Stereolabs provides a very comprehensive API to obtain depth information. Therefore, after YOLOX outputs the 2D coordinates of the horn antenna in the image plane, we only need to call the corresponding interface function of the ZED II camera to obtain the 3D coordinates of the horn antenna with respect to RIS.

The YOLOX algorithm and the stereo matching algorithm both are running on the personal computer (PC). ZED II camera sends the environmental photos to PC through the USB interface, and then PC sends the calculated horn antenna’s location information to RIS control board through the serial port. It is tested that the whole process from acquiring photos to outputting horn antenna’s location takes about 85 ms. Moreover, to ensure detection accuracy, the YOLOX-DarkNet53 model is used in the designed vision module. However, a lighter version can be used to obtain faster inference, e.g., YOLOX-Tiny [28].

B. RIS and Codebook Design

The RIS in Fig. 5 contains 400 \((20 \times 20)\) units with a quarter wavelength spacing between the individual units. Meanwhile, the units are square structure with quarter wavelength sides, as shown in Fig. 6. Each unit consists of a microwave structure on the top layer, a metal plate, a bias line and a PIN diode [35]. The bias voltage across the PIN diode can be changed to control the PIN diode states (forward bias state or reverse bias state). Here we denote “1” state as forward bias state and “0” as the reverse bias state. For different operating frequencies, the unit has different modulation phases. The modulation phase difference between “0” state and “1” state should be 180 degree. However, PIN diodes with different batches have different equivalent circuit parameters, and thus the most suitable operating frequency of the unit is decided based on the test results. The modulation phase of the RIS unit varies as a function of operating frequency as shown in Fig. 6.

It is seen that the optimal operating frequency is 5.4 GHz. Therefore, the designed RIS board is a square board with sides length of 28 cm. Generally, the specific modulation phase values corresponding to the two states are not determined. For the convenience of the codebook design, we assume that the modulation phase of the unit controlled by the PIN diode in “1” state is \(-\pi/2\), and the modulation phase of the unit controlled by the PIN diode in the “0” state is \(\pi/2\).

According to (9), in order to calculate the optimal modulation phase \(\alpha_{mn}^{\theta_{rx}, \varphi_{rx}}\), we should obtain the initial phase \(\beta_{mn}\) of the incident wave in advance. The way to calculate the phase \(\beta_{mn}\) is not fixed and can be different for two cases: (I) RIS works in near-field as a passive array antenna of the BS; (II) RIS works in far-field to assist the communications between the BS and the UE.

Since the maximum aperture of the RIS is 20\(d\), the cut-off distance for the RIS to distinguish the near field from the far field is 800\(d^2/\lambda\). In case I, the distance between the transmitting antenna and the RIS is less than 800\(d^2/\lambda\), i.e., \(\sqrt{x_{tx}^2 + y_{tx}^2 + z_{tx}^2} < 800d^2/\lambda\). Thus we need to consider the distance between the transmitting antenna and each unit separately. We assume that the center of the transmitting antenna is facing the center of the RIS and denote the vertical distance between the transmitting antenna and the RIS as \(d_{feed}\), i.e., the 3D coordinates of the BS with respect to RIS is \((d_{feed}, 0, 0)\). The incident wave phase \(\beta_{mn}^t\) of the \(mn\)-th unit can be calculated as

\[
\beta_{mn}^t = \frac{2\pi}{\lambda} d_{feed} - \frac{2\pi}{\lambda} \sqrt{d^2 \left(\frac{M + 1}{2} - m\right)^2 + d^2 \left(n - \frac{1 + N}{2}\right)^2 + d_{feed}^2}. \tag{14}
\]

In case II, the distance between the transmitting antenna and the RIS is greater than 800\(d^2/\lambda\), i.e., \(\sqrt{x_{tx}^2 + y_{tx}^2 + z_{tx}^2} > 800d^2/\lambda\). Thus the incident electromagnetic waves can be...
approximated as plane waves with respect to the RIS. We assume that the pitch and azimuth angles of the transmitting antenna with respect to the RIS are $\theta_{tx}$ and $\varphi_{tx}$, respectively. According to (12), (13), and BS’s 3D coordinates $(x_{tx}, y_{tx}, z_{tx})$, we can obtain $\theta_{tx}$ and $\varphi_{tx}$. Then we can calculate the incident phase $\beta_{mn}^{II}$ of the unit as

$$\beta_{mn}^{II} = \frac{2\pi d}{\lambda} \left( \frac{M + 1}{2} - m \cos \theta_{tx} \right) + \frac{2\pi d}{\lambda} \left( n - \frac{1 + N}{2} \sin \theta_{tx} \sin \varphi_{tx} \right). \quad (15)$$

The optimal modulation phase $\alpha_{mn}^{\theta_{rx}, \varphi_{rx}}$ can be obtained by substituting (14) or (15) into (9). However, the individual units can only provide two different phases ($\pi/2$ or $-\pi/2$), and thus we need to quantize $\alpha_{mn}^{\theta_{rx}, \varphi_{rx}}$ by one bit. The specific scheme of the quantization is

$$\alpha_{mn, \text{quan}}^{\theta_{rx}, \varphi_{rx}} = \begin{cases} \frac{\pi}{2}, & \alpha_{mn}^{\theta_{rx}, \varphi_{rx}} \in [0, \pi), \\ -\frac{\pi}{2}, & \alpha_{mn}^{\theta_{rx}, \varphi_{rx}} \in [-\pi, 0). \end{cases} \quad (16)$$

According to (9), (14), (15), and (16), we can arbitrarily set the desired outgoing direction $(\theta_{rx}, \varphi_{rx})$ to get the corresponding codebook, e.g., under case I we can calculate the RIS codeword when the desired pitch angle $\theta_{rx}$ is 90° and the azimuth angle $\varphi_{rx}$ is 10°, as shown in Fig. 7. The “0” in Fig. 7 represents that the PIN diode is in reverse bias state and the modulation phase of the controlled unit is $\pi/2$. The “1” in Fig. 7 represents that the PIN diode is in forward bias state and the modulation phase of the controlled unit is $-\pi/2$.

Generally, the location of the BS is fixed and the value of $\beta_{mn}$ can be determined in advance. Thus the codebook can be pre-calculated and solidified.

### C. Control Board Design

In order to meet the real-time beam tracking in the mobile scene, the vision-aided RIS needs to have a fairly fast beam switching speed, which imposes high requirements on the code switching speed. Meanwhile, the general shift register scheme [16] can not meet the demand of real-time codeword refreshing, which limits the advantages of visual assistance. Therefore, we cascaded two low-cost, high-I/O-density FPGA chips (Intel Cyclone IV EP4CE15F23C8N) in the control board, which can easily run at up to 200 MHz system clock frequency and control each PIN diode individually. The codeword switching speed is much faster than the general shift register scheme.
Fig. 8. Schematic diagram of RIS High-Speed control board. The primary FPGA chip can receive codewords or external instructions through the external interface and can control the secondary FPGA or transmit codewords through the AXI-stream-4 bus. The control board can control the state of each RIS unit independently through the I/O pins of the FPGAs.

Fig. 8 shows the schematic diagram of the RIS control circuit, including a primary FPGA controller, a secondary FPGA controller, an external communications interface, bias circuits, and power management module. The primary FPGA is responsible for controlling the first 10 rows (200 in total) of PIN diodes and external communications, while the secondary FPGA is responsible for controlling the other 200 PIN diodes. The primary FPGA and the secondary FPGA are connected on the printed circuit board through a parallel 16-bit AXI-Stream-4 bus, where the signal line consists of 16-bit-tdata, 1-bit-tready, 1-bit-tvalid, and 1-bit-tclk. The bus follows the AXI-Stream-4 protocol for communications. The cascading scheme can support more FPGA chips working together, and thus is more easily reusable for RIS with a larger number of units. The reference clock rate of 100 MHz allows for high-speed data transfer from primary to secondary FPGA chips, which can reach a peak speed of 1.6 Gbps.

In the designed bias circuits, each I/O of the FPGA chips is connected in series with PIN diodes, current limiting resistors and a 1.1v voltage source. Each I/O can output 3.3v or 0v voltage (1 or 0) to provide each PIN diode with 1.5v forward bias or 1.1v reverse bias and to limit the current to 5 mA. The forward bias state of the PIN diode corresponds to “1” in the codeword, and the reverse bias state corresponds to “0”. In addition, the power management part is designed with an integrated 4-channel DC-DC chip LTM4644IY, which generates the 1.2v, 2.5v, 3.3v power rails required by the FPGA chips, and the 1.1v reverse bias voltage required by the PIN diode. The 400 control I/O signals are connected to the RIS front panel via eight 2.54 mm connectors. Fig. 8 shows the finished control board. Since the current passing through each PIN diode is very small, the total power consumption of the RIS including the bias circuit and FPGA is less than 0.5 W.

In order to enhance the robustness and the universality of the prototype system, the control board has three working modes: Index control mode: In this mode, the pre-calculated codebook needs to be downloaded to the Flash memory of the FPGA in advance. The PC selects the appropriate codeword to be switched and transmits the corresponding index number to the primary FPGA through the external communications interface (serial port or SPI). Then the primary FPGA sends the index number to the secondary FPGA through the parallel AXI-Stream-4 bus. Dynamic codebook mode: In this mode, the codewords are calculated on the PC in real time. Then the codeword stream should be input through the external communications interface, and the primary FPGA sends codeword to the secondary FPGA through the parallel AXI-Stream-4 bus. Codebook download mode: After power-on, the FPGA chips rewrite the preset codebook through the external communications interface, and then the primary FPGA sends the half of the codebook to the secondary FPGA through the parallel AXI-Stream-4 bus.

D. Peer-to-Peer Communications System Design

To validate the effectiveness of the vision-aided RIS, we design a peer-to-peer communications system to simulate the physical link between the BS and the UE, as shown in Fig. 10. The RFSoc development board, the transmitting RF chain, and the transmitting horn antenna are the main components of the transmitter, which emulates the BS. The RFSoc development board, the receiving RF chain, and the receiving horn antenna are the main components of the receiver, which emulates the UE. In the transmitter, the business data is transmitted from PC through the ARM processor of the RFSoc chip through User Datagram Protocol (UDP), which runs the LwIP protocol stack to parse the UDP packets. In addition, there...
is a simple MAC program to package the data into physical layer frames, and then the ARM processor sends the data to the programmable logic part via XDMA and AXI-S bus. Modulation and demodulation Verilog programs are running in the programmable logic part to process digital baseband signals in real time. Meanwhile, the receiver demodulates the bit stream and transmits it to the host via UDP. The digital baseband can also upload the intermediate information generated in the demodulation to the PC and display the intermediate information, such as constellation diagram, channel status information, etc.

The baseband frame format in the peer-to-peer communications system is based on the 20Mhz version of 802.11n standard [36]. However, in order to take full advantage of the RIS’s operating bandwidth and to increase the communication rate, we increase the bandwidth to 40Mhz. Thus the whole baseband contains 52 data subcarriers and 4 pilot subcarriers with a subcarrier spacing of 625 kHz. The baseband algorithm is divided into two parts, i.e., the transmitter and the receiver, to be implemented separately. The baseband algorithm of the transmitter mainly includes seven steps: scrambling, convolutional coding, interleaving, quadrature amplitude modulation (QAM), insertion of pilot, inverse discrete fourier transform (IDFT), and insertion of cyclic prefix (CP). At the receiver side, the wireless signal is demodulated by the baseband algorithm to obtain the transmitted data after down-conversion and decimation filtering. The receiver’s baseband algorithm consists of ten steps: symbol synchronization, carrier frequency offset (CFO) estimation and compensation, CP removal, DFT, channel estimation, channel equalization, symbol demodulation, deinterleaving, decoding, and descrambling.

Next, we will provide a detailed description of the processing of the RF link. After the 40 MHz baseband I/Q signal is generated by the digital transmit baseband, the baseband signal is interpolated to the sampling rate of 4.8 GHz, and then is mixed with a 48-bit digital oscillator (NCO). The baseband signal is placed at 5.4 GHz on the spectrum via digital up-conversion (DUC) and is sent to the RF-DAC. The RF-DAC supports bandpass sampling and is set to the third Nyquist zone mode, which can optimize the signal power and in-band flatness in the third zones. Since the signal generated by RF-DAC has image signals and the second harmonics, the transmit RF chain is configured with a 5.2 GHz-5.8 GHz band-pass filter to remove the unwanted spurious signals. After filtering, the RF signal passes through the driver amplifier and power amplifier, and is then transmitted through the horn antenna. Similarly, in the transmitter, the receiving RF chain is configured with a low-noise amplifier and a band-pass filter. The RF-ADC has a sampling rate of 4.8 GHz and works in the third Nyquist zone. The sampled signal is decimation filtered and digital down-converted (DDC) to obtain the baseband I/Q signal. The signal processing flow of the peer-to-peer communication system is shown in Fig. 11.

After testing, the peak transmission rate of the system can reach 52.4 Mbps on 16QAM modulation, which can transmit 2K high-definition video encoded by H.264 in real time. Additionally, the system can display the demodulation constellation diagram in real time. The parameters of the peer-to-peer communications system are shown in Table I, which meets the verification requirements of the vision-aided RIS.

### V. Experimental Results

In this section, we simulate the radiation pattern of the RIS and compare the simulation results with the radiation patterns of the actual test. Then we test the codeword refresh speed of the control board and examine the beam tracking effect of the vision-aided RIS in two classical cases.

#### A. Radiation Pattern of RIS

Taking case I where the RIS works in near-field condition as an example, we assume that the transmitting antenna is located at three electromagnetic wave wavelengths directly in front of the RIS. According to (9), (14) and (16), we fix the desired pitch angle $\theta_{rx}$ as 90° and the desired azimuth angle $\phi_{rx}$ as $-40^\circ$, $-30^\circ$, $-20^\circ$, $-10^\circ$, 0°, 10°, 20°, 30°, 40°, respectively, to get the corresponding codewords. Then we use the three-dimensional electromagnetic field simulation software CST to calculate the radiation pattern of the RIS as shown in Fig. 13(a). As can be seen from the figure, the angles of the centers of the main lobes of radiation patterns are consistent with the desired angles, which verifies the correctness of the codebook design principle. Meanwhile, we test the practical radiation patterns of the RIS in an anechoic chamber.
free of other electromagnetic wave interference, as shown in Fig. 12. The test tool is the vector network analyzer and the antenna test turntable system. Based on the previously calculated codewords, we obtain the practical radiation patterns of multiple reception angles as shown in Fig. 13(b). The test results are basically consistent with the simulation, which further proves the correctness and the feasibility of the proposed codebook design.

B. Control Board Test

In the proposed prototype system, in order to ensure the codeword refresh speed, the control board uses two FPGA chips to control each diode on the RIS independently. Here we use an oscilloscope to test the codeword switching time. The first signal connected to the oscilloscope is the control signal sent by the PC through the serial port, and the other signal is the bias signal on the diode. The test results are shown in Fig. 14, from which it can be seen that the time from the control signal sent to the completion of the codeword refresh is about 85 us.

C. Test Under Case I

In case I, the RIS works as a passive array antenna at BS to achieve beamforming. In the test scenario, we use the transmitting horn antenna located at a distance of 3 electromagnetic wavelengths in front of the RIS as the feeding antenna of the BS and use the receiving horn antenna at a distance of 2.2 meters from the RIS as the UE. The layout of the test scenario is shown in Fig. 15. The task of the vision-aided RIS is to accurately reflect the electromagnetic waves emitted by the feeding antenna to the horn antenna at the receiver side based on the visual information. In case I, the RIS works in near-field condition and the location of the transmitting antenna is fixed and known, and thus we can calculate the RIS codebook in advance according to (9), (14) and (16). As for the vision-based beam tracking scheme, we use object detection and stereo vision technology to obtain the 3D coordinates of the horn antennas with respect to the RIS. Then beam tracking is implemented in two stages. In the first stage, we can distinguish the BS from the UE easily due to the fixed location of the BS. Thus we can determine the codebook to be used. In the second stage, according to UE’s 3D coordinates, RIS can obtain the desired beam direction and select the optimal codeword among the codebook to achieve beam tracking. We can obtain the UE’s prediction bounding box and the UE’s direction output by vision algorithm at a certain moment as shown in Fig. 16.

During the test, we move the UE back and forth at a fixed angular velocity 28°/s within the coverage of the RIS beamforming. In addition, we set the initial codeword of the RIS as the codeword with the desired pitch angle \( \theta_{rx} \) of 90° and the desired azimuth angle \( \varphi_{rx} \) of 0°. To highlight the effectiveness of the vision, we conduct two experiments both with and without visual assistance. When there is no visual assistance, the traditional beam-sweeping method is used to achieve beam tracking [17], where the feedback procedure is ignored and the UE side is directly connected to the RIS. The specific beam-sweeping strategy is: firstly, the full-range beam-sweeping is used to find the optimal beam direction; then during the beam tracking process, if the signal-to-noise ratio (SNR) of the UE side is 6 dB lower than the maximum...
Fig. 13. Radiation patterns of RIS. The desired pitch angle $\theta_{rx}$ of the reflection beam is fixed as $90^\circ$ and the desired azimuth angle $\phi_{rx}$ is set as $-40^\circ$, $-30^\circ$, $-20^\circ$, $-10^\circ$, $0^\circ$, $10^\circ$, $20^\circ$, $30^\circ$, $40^\circ$, respectively. Nine RIS radiation patterns are simulated or tested. (a) Simulation results of RIS radiation patterns. (b) Practical test results of RIS radiation patterns.

SNR, a small range of scanning is performed near the current beam direction.

We calculate the real-time SNR at the UE to evaluate the beam tracking performance of the vision-aided RIS. The SNR variation curves are displayed in Fig. 17. It is seen that when the UE moves around, the SNR curve fluctuates steadily up and down around 35 dB with visual assistance, which ensures the high quality of the communications. Meanwhile, there is no additional feedback process and beam training overhead during the whole communication process. However, if the conventional beam-sweeping method is adopted, there will be a significant drop in SNR at some moments, e.g., the interval of 9200 ms to 11500 ms in the Fig. 17. When the received SNR is poor, the RIS needs to perform beam sweeping to achieve beam tracking. However, communications has to stand by during the beam-sweeping process, which introduces a certain beam training overhead.

D. Test Under Case II

In case II, the RIS is used as an independent component to assist the communications between the BS and the UE. Here we use a transmitting horn antenna at a distance of 3 meters from the RIS as the BS, and use a receiving horn antenna at a distance of 2.2 meters from the RIS as the UE. Meanwhile, the LOS path between BS and UE is blocked. The layout of the test scenario is shown in Fig. 19. When the LOS path is blocked, the received SNR on the UE side is poor, which reduces the spectral efficiency significantly. The task
for vision-aided RIS is to create another communications path between BS and UE to improve the SNR based on the visual information. In case II, the RIS works in far-field conditions, and thus the codebook needs to be redesigned according to (9), (15), and (16) in Section IV. Similar to the beam tracking scheme in case I, the vision-aided RIS needs to find the BS and calculate the exact UE's direction with respect to the RIS, and then adjusts the reflection coefficients of the RIS according to the codebook timely. We assume that the UE moves back and forth at a fixed angular velocity $28^\circ/s$ in the region where the LOS path is blocked and the initial codeword of the RIS is with the desired pitch angle $\theta_{rx}$ of $90^\circ$ and the desired azimuth angle $\varphi_{rx}$ of $0^\circ$. Correspondingly, we conduct the communications test in both the situation with and without visual assistance. Similarly, in the absence of visual assistance, we use the beam-sweeping method to realize beam tracking and the corresponding SNR variation curves are plotted in Fig. 19. It can be seen that even if the LOS path between the BS and the UE is blocked, the vision-aided RIS can adjust another direct path to compensate for the performance degradation, keeping the SNR stable between 20 dB and 25 dB without additional beam training overhead and feedback overhead. However, without the help of visual information, the SNR can fall below 15 dB at some moments, which raises the communications delay.

VI. DISCUSSION AND LIMITATIONS

In the proposed prototype system, thanks to the high speed control board, the time to refresh the RIS codeword does not exceed 100 us, and thus the delay of beam tracking depends entirely on the calculation time of the UE's 3D coordinates (about 85 ms). Meanwhile, it is can be seen from Fig. 13 that the main lobe width of the RIS beam is greater than 10$^\circ$. Thus the proposed vision-aided RIS enables the real-time beam tracking as long as the UE’s angular velocity does not exceed 59 $^\circ/s$, which is much higher than the angular velocity of people or cars in real environment. Specifically, for indoor communications, assuming the distance between human and RIS is 2 m, the prototype system can work well as long as the speed of human does not exceed 2 m/s; while for outdoor communications, assuming the distance between car and RIS is 20 m, the prototype system can work well as long as the speed of car does not exceed 72 km/h. In addition, the vision algorithms used in this prototype are unoptimized versions, and lightweight neural network models, e.g., YOLOX-Tiny, can be used to make beam tracking performance better. Therefore, the proposed vision-aided RIS in this paper could be widely applicable to various communications systems. However, the proposed prototype system still has some limitations, which are divided into two aspects. The first one is that the transmitter and receiver are equipped with horn antennas due to the limited size of the RIS, and only a single mobile UE is considered. When the size of the RIS board is large enough, the proposed beam tracking scheme is expandable to the situation of ordinary antenna, e.g., omni-directional antenna. As for the case of multi-antenna UE as well as multiple UEs, suitable codewords and proper communication scheme need to be designed, which are the future research directions of the vision-aided RIS.

The second limitation is imposed by the vision module itself. Introducing a vision module into the prototype system not only brings additional costs, but also makes the system ineffective in extreme situations, such as in the dark or when
the BS/UE is obscured. In fact, with the maturity of edge computing, the actual cost of implementing the vision module in RIS system can be less than one-fifth of the cost of RIS itself and is ignorable compared to a regular RF chain. With regard to the extreme situations, we can combine other out-of-band sensors, e.g., LIDAR or infrared sensor, to detect the BS/UE or return to the mode of beam-sweeping to detect BS/UE. Adopting multimodal methods to make up for the lack of vision is also one of the future research directions.

VII. CONCLUSION

In this paper, we propose a novel computer vision-based approach to aid RIS for dynamic beam tracking and implement the corresponding prototype verification system. A camera is attached at the RIS to obtain the visual information about the surroundings, with which RIS identifies the incident beam direction and the desired reflected beam direction and then adjusts the reflection coefficients according to the pre-designed codebooks. Next we build a 20-by-20 RIS running at 5.4 GHz and develop a high-speed control board to ensure the real-time refresh of the reflection coefficients. Meanwhile we implement an independent peer-to-peer communications system to simulate the physical link between the BS and the UE by referring to the 802.11n physical layer standard. The prototype system not only enables beam training overhead and feedback links, but also overcomes the challenges of real-time refreshing and storage for codewords.

We calculate the radiation patterns by using the three-dimensional electromagnetic field simulation software CST and then compare the simulation results with the practical test results in an anechoic chamber. Both simulation and test results of the radiation patterns show that the angle of the main lobe center of the radiation pattern is consistent with the desired angle, which verifies the correctness of the codebook design principle. Then we test the vision-aided RIS prototype system in two cases and compare it with the traditional beam-sweeping method. In case I, the RIS works as a passive array antenna at BS to achieve beamforming. In case II, the RIS is used as an independent component to assist the communications between the BS and the UE. Both experimental results show that the vision-aided RIS can achieve real-time beam tracking and stabilize the received SNR of the UE.

REFERENCES

[1] S. Kisseleff, W. A. Martins, H. Al-Hraishawi, S. Chatzinotas, and B. Ottersten, “Reconfigurable intelligent surfaces for smart cities: Research challenges and opportunities,” IEEE Open J. Commun. Soc., vol. 1, pp. 1781–1797, 2020.
[2] R. Liu, Q. Wu, M. Di Renzo, and Y. Yuan, “A path to smart radio environments: An industrial viewpoint on reconfigurable intelligent surfaces,” IEEE Wireless Commun., vol. 29, no. 1, pp. 202–208, Feb. 2022.
[3] W. Long, R. Chen, M. Moretti, W. Zhang, and J. Li, “A promising technology for 6G wireless networks: Intelligent reflecting surface,” J. Commun. Inf. Netw., vol. 6, no. 1, pp. 1–16, Mar. 2021.
[4] M. A. El Mossallamy, H. Zhang, L. Song, K. G. Seddik, Z. Han, and G. Y. Li, “Reconfigurable intelligent surfaces for wireless communications: Principles, challenges, and opportunities,” IEEE Trans. Cogn. Commun. Netw., vol. 6, no. 3, pp. 990–1002, Sep. 2020.
[5] Y. Liu et al., “Reconfigurable intelligent surfaces: Principles and opportunities,” IEEE Commun. Survey Tuts., vol. 23, no. 3, pp. 1546–1577, 3rd Quart., 2021.


**H. Yang et al.** “A programmable metasurface with dynamic polarization, scattering and focusing control,” Sci. Rep., vol. 6, no. 1, p. 35692, 2016.

**Z. Ge, S. Liu, F. Wang, Z. Li, and J. Sun.** “YOLOX: Exceeding YOLO series in 2021,” 2021, arXiv:2107.08430.

**A. Bochkovskiy, C.-Y. Wang, and H.-Y. Mark Liao.** “YOLOv4: Optimal speed and accuracy of object detection,” 2020, arXiv:2004.10934.

**H. Zhang, M. Cissé, Y. N. Dauphin, and D. Lopez-Paz.** “mixup: Beyond empirical risk minimization,” in Proc. Int. Conf. Learn. Represent. (ICLR), Apr. 2018, pp. 1–13.

**J. Redmon and A. Farhadi.** “YOLOv3: An incremental improvement,” 2018, arXiv:1804.02767.

**H. Hirschmuller.** “Stereo processing by semiglobal matching and mutual information,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no. 2, pp. 328–341, Dec. 2008.

**H. Laga, L. V. Jospin, F. Boussaid, and M. Bennamoun.** “A survey on deep learning techniques for stereo-based depth estimation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 44, no. 4, pp. 1738–1764, Apr. 2022.

**R. Hartley and A. Zisserman.** *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.

**P. Li, T. Yang, J. Ren, and Y. Yin.** “Design of 1-bit reconfigurable reflectarray based on miniaturized reconfigurable unit,” in IEEE MTT-S Int. Microw. Symp. Dig., Nov. 2021, pp. 370–372.

**E. Perahia and R. Stacey.** *Next Generation Wireless LANs: Throughput, Robustness, and Reliability in 802.11n*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

**Ming Ouyang** (Graduate Student Member, IEEE) received the B.Eng. degree from Sichuan University, Chengdu, China, in 2021. He is currently pursuing the M.S. degree with the Department of Automation, Tsinghua University, Beijing, China. His research interests include MIMO-OFDM baseband processing, reconfigurable intelligent surface (RIS), design and implementation of wireless communication prototype systems, and orthogonal frequency division multiplexing (OFDM) modulation.

**Feifei Gao** (Fellow, IEEE) received the B.Eng. degree from Xi’an Jiaotong University, Xi’an, China, in 2002, the M.Sc. degree from McMaster University, Hamilton, ON, Canada, in 2004, and the Ph.D. degree from the National University of Singapore, Singapore, in 2007. Since 2011, he has been with the Department of Automation, Tsinghua University, Beijing, China, where he is currently an Associate Professor. He has authored/coauthored more than 150 refereed IEEE journal articles and more than 150 IEEE conference proceeding papers that are cited more than 15000 times in Google Scholar. His research interests include signal processing for communications, array signal processing, convex optimizations, and artificial intelligence assisted communications. He served as the Symposium Co-Chair for the 2019 IEEE Conference on Communications (ICC), the 2018 IEEE Vehicular Technology Conference Spring (VTC), the 2015 IEEE ICC, the 2014 IEEE Global Communications Conference (GLOBECOM), and the 2014 IEEE Vehicular Technology Conference Fall (VTC), and a technical committee member for more than 50 IEEE conferences. He served as an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, IEEE WIRELESS COMMUNICATIONS LETTERS, and China Communications, the Lead Guest Editor for IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and IEEE TRANSACTIONS ON COGNITIVE COMMUNICATIONS AND NETWORKING, and the Senior Editor for IEEE SIGNAL PROCESSING LETTERS and IEEE COMMUNICATIONS LETTERS.

**Yucong Wang** (Graduate Student Member, IEEE) received the B.Eng. degree from Xidian University, Xi’an, China, in 2021. He is currently pursuing the Ph.D. degree with the Department of Automation, Tsinghua University, Beijing, China. His research interests include ultra-wideband communications, THz communication systems, radar systems, and ISAC.

**Shun Zhang** (Senior Member, IEEE) received the B.S. degree in communication engineering from Shandong University, Jinan, China, in 2007, and the Ph.D. degree in communications and signal processing from Xidian University, Xi’an, China, in 2013. He is currently with the State Key Laboratory of Integrated Services Networks, Xidian University, where he is also a Professor. He is an inventor of 16 granted patents, including a PCT patent authorized by the U.S. Patent and Trademark Office. He has authored or coauthored more than 80 journals and conference papers. His research interests include massive MIMO, millimeter wave systems, RIS assisted communications, deep learning for communication systems, orthogonal time frequency space (OTFS) systems, and multiple access techniques. He received the two best paper awards in conferences and the two prize awards in natural sciences for research excellence from the China Institute of Communications and the Chinese Institute of Electronics. He is an Editor of *Physical Communication*.

**Puchu Li** was born in Shanshi, China. He received the B.Sc. degree in electronic engineering from Xidian University, Xi’an, China, in 2020, where he is currently pursuing the master’s degree in electromagnetic field and microwave technology with the National Key Laboratory of Antennas and Microwave Technology.

**Jian Ren** (Member, IEEE) was born in Shandong, China. He received the B.Sc. and M.Eng. degrees in electronic engineering from Xidian University, Xi’an, China, in 2012 and 2015, respectively, and the Ph.D. degree in electronic engineering from the City University of Hong Kong in 2018. From March 2015 to August 2015, he was a Research Assistant with the City University of Hong Kong. In 2019, he joined the National Key Laboratory of Antennas and Microwave Technology, Xidian University, where he is currently an Associate Professor. He has authored or coauthored more than 60 refereed journal articles. His current research interests include dielectric resonator antenna, millimeter wave antennas, and metamaterials. He received the Honorable Mention at the Student Best Paper Competition from the 2018 IEEE 7th Asia-Pacific Conference on Antennas and Propagation (APCAP). He has served as the TPC Co-Chair for the 2021 Cross Strait Radio Science and Wireless Technology Conference (CSRSWTC). He has served as a Reviewer for different peer-reviewed journals, including the IEEE TRANSACTIONS ON ANTENNAS AND PROPAGATION, IEEE TRANSACTIONS ON MICROWAVE THEORY AND TECHNIQUES, *Applied Physics Letters, Optics Express*, and *Optics Letters*.

---

**Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.**