

Integrated Imaging and Communication with Reconfigurable Intelligent Surfaces

Hao Luo and Ahmed Alkhateeb

Abstract—Reconfigurable intelligent surfaces, with their large number of antennas, offer an interesting opportunity for high spatial-resolution imaging. In this paper, we propose a novel RIS-aided integrated imaging and communication system that can reduce the RIS beam training overhead for communication by leveraging the imaging of the surrounding environment. In particular, using the RIS as a wireless imaging device, our system constructs the scene depth map of the environment, including the mobile user. Then, we develop a user detection algorithm that subtracts the background and extracts the mobile user attributes from the depth map. These attributes are then utilized to design the RIS interaction vector and the beam selection strategy with low overhead. Simulation results show that the proposed approach can achieve comparable beamforming gain to the optimal/exhaustive beam selection solution while requiring 1000 times less beam training overhead.

I. INTRODUCTION

Integrated sensing and communication (ISAC) [1] has been identified as a key feature of future wireless systems. By incorporating these two functions into a single system, ISAC has the potential to enhance spectrum efficiency and reduce hardware cost and power consumption. Furthermore, ISAC has the potential to benefit both functionalities through communication-aided sensing and sensing-aided communication. Meanwhile, reconfigurable intelligent surfaces (RISs) have emerged as a promising approach to extend coverage and overcome blockages in both communication and sensing systems. The RIS can steer incident signals toward desired directions by adjusting the phase shifts of the passive reflecting elements. However, in RIS-assisted communication, the optimal configuration of the reflecting elements requires significant beam training overhead. Therefore, in this paper, we aim to develop an RIS-aided integrated imaging and communication system, where the RIS-based imaging of the surrounding environment can be used to mitigate the communication beam training overhead.

Several studies have explored the use of RISs in ISAC systems [2], [3]. In [2], the RIS-aided dual-function radar and communication system is proposed, where the transmit precoding and passive reflection matrices are jointly optimized. In [3], the authors study the joint design of the active beamforming of radar and the passive reflection matrices of two RISs in the communication radar coexistence system. So far, prior work has mainly focused on the interplay between communication and target detection, while other wireless sensing functionalities, e.g., imaging, have not been widely investigated.

The authors are with Arizona State University (Email: h.luo, alkhateeb@asu.edu). This work is supported by the National Science Foundation under Grant No. 2229530.

In this paper, we propose an RIS-aided integrated imaging and communication system, where the imaging-aided communication is achieved. Specifically, the contributions of this paper are organized as follows:

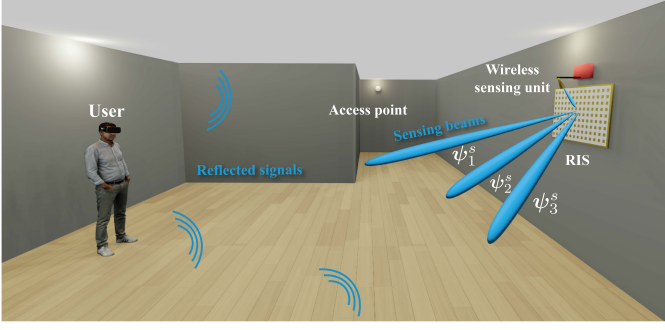
- We introduce a novel RIS-aided integrated imaging and communication system. The system leverages the high spatial dimensions of the RIS to perform wireless imaging and build a depth map of the surrounding environment. This depth map is then used to design the RIS interaction vector for communication with low overhead.
- We propose a user detection algorithm to extract the user position from an estimated depth map. Using the user position, we design the RIS interaction vector for communication. Then, we develop a beam selection strategy by considering a pre-defined RIS interaction codebook.

The simulation results demonstrate the capability of the designed solutions in achieving high beamforming gain and significantly reducing the RIS beam training overhead. This highlights the potential of the proposed RIS-aided integrated imaging and communication system.

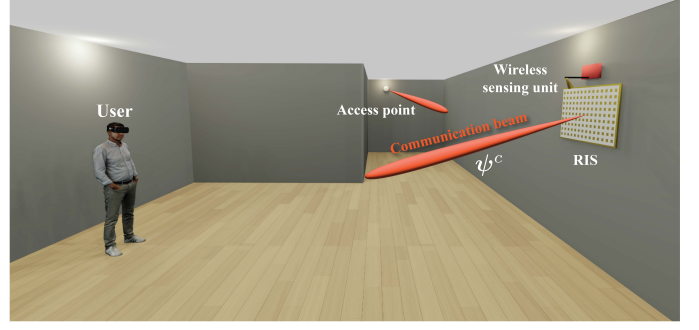
Notation: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar. $\mathcal{A}$ and $\mathcal{A}$ are sets of scalars and vectors. $\mathbf{A}^T$, $\mathbf{A}^H$, and $\mathbf{A}^*$ are the transpose, Hermitian (conjugate transpose) and conjugate of $\mathbf{A}$. $[\mathbf{a}]_n$ is the n^{th} element of the vector $\mathbf{a}$. $\text{diag}(\mathbf{a})$ is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal. $\mathbf{A} \odot \mathbf{B}$ is the Hadamard product of $\mathbf{A}$ and $\mathbf{B}$. $\mathcal{N}(m, R)$ is a complex Gaussian random variable with mean m and covariance R . $\text{Re}(z)$, $\text{Im}(z)$, and $\arg(z)$ are the real part, the imaginary part, and the phase angle of the complex number z . $f(t) * g(t)$ is the continuous-time convolution of two signals $f(t)$ and $g(t)$.

II. SYSTEM MODEL

As illustrated in Fig. 1, we consider an RIS-aided integrated imaging and communication system, which consists of a mmWave access point (AP), a user, an RIS, and a mmWave wireless sensing unit. The mmWave access point, functioning as a transmitter, communicates with the user via the RIS. For simplicity, both the access point and the user equipment (UE) are assumed to have a single antenna structure. Also, it is assumed that there is no direct link between the access point and the user equipment. The mmWave wireless sensing unit is placed near the RIS and illuminates the non-line-of-sight (NLoS) area of the access point through the RIS. Following the design proposed in [4], the wireless sensing unit comprises a transmitter and a receiver that are connected to a shared single antenna through a self-isolation circuitry. The shared single



(a) Imaging Stage



(b) Communication Stage

Fig. 1. This figure shows the integrated imaging and communication system. In the imaging/sensing stage, the wireless sensing unit transmits the sensing signals to the RIS via a feeding antenna. The sensing signals are reflected towards the environment by the RIS, which then reflects the backscattered/reflected signals back to the wireless sensing unit. The received signals are processed by the wireless sensing unit to construct a depth map of the environment, which enables the system to design the RIS interaction vector for communication.

antenna serves as a feeding antenna and transmits sensing signals to the RIS in order to achieve high spatial resolution.

The RIS is assumed to be equipped with N reconfigurable elements, and each element is modeled as a phase shifter. Denote the RIS interaction matrix for communication by $\Psi^c = \text{diag}(\psi^c) \in \mathbb{C}^{N \times N}$, where $\psi^c = [e^{j\phi_1}, \dots, e^{j\phi_N}]^T$ is the interaction vector with unit modulus entries. The interaction matrix for sensing Ψ^s can be defined similarly. In this paper, our objective is to detect the NLoS user from an estimated depth map and utilize it to devise the RIS interaction vector for communication. It is worth noting that the RIS-aided wireless sensing for depth estimation can be well-performed by an existing approach in the literature [4]. Besides, we assume that the user is static during sensing and communication, and the scenario with the moving user is left to future work. Next, the communication and sensing models are described.

A. Communication Model

We assume the access point sends a complex symbol x_c with the average power constraint $\mathbb{E}[|x_c|^2] = \mathcal{E}_c$, and the RIS reflects the incident signal with the interaction vector ψ^c . Then, the received signal can be expressed as

$$y_c = \mathbf{h}_R^T \Psi^c \mathbf{h}_T x_c + w_c = (\mathbf{h}_R \odot \mathbf{h}_T)^T \psi^c x_c + w_c, \quad (1)$$

where $\mathbf{h}_T \in \mathbb{C}^N$ is the channel between the AP and the RIS, and $\mathbf{h}_R \in \mathbb{C}^N$ is the channel between the RIS and the user. $w_c \sim \mathcal{N}(0, \sigma_w^2) \in \mathbb{C}$ is the received noise. The channel $\mathbf{h}_T$ can be defined as

$$\mathbf{h}_T = \sum_{\ell=1}^L \alpha_\ell \mathbf{a}(\theta_\ell^{\text{az}}, \theta_\ell^{\text{ze}}), \quad (2)$$

where α_ℓ is the complex-valued channel gain of path ℓ and $\mathbf{a}(\cdot)$ is the far-field transmit/receive array response vector of the RIS. $\theta_\ell^{\text{az}}, \theta_\ell^{\text{ze}}$ denote the azimuth and zenith angles of arrival, relative to the RIS. The channel $\mathbf{h}_R$ can be defined similarly as $\mathbf{h}_T$.

B. Sensing Model

For the sensing model, we adopt a wideband FMCW radar transceiver with a complex-baseband architecture [5], [6]. The FMCW transmit signal is referred to as a radar frame, which contains a sequence of M_{chirp} chirps with a repetition interval of T_{PRI} . A single linear chirp signal $a_{\text{BP}}(t)$ can be written as

$$a_{\text{BP}}(t) = \begin{cases} \cos(2\pi f_0 t + \pi S t^2) & 0 \leq t \leq T_{\text{active}}, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

where f_0 is the starting chirp frequency, and T_{active} is the duration of the chirp signal. $S = \text{BW}/T_{\text{active}}$ is the slope of the linear chirp with the bandwidth BW. Then, the transmit signal of a radar frame $x_s^{\text{BP}}(t)$ can be formulated as

$$x_s^{\text{BP}}(t) = \sqrt{\mathcal{E}_s} \sum_{v=0}^{M_{\text{chirp}}-1} a_{\text{BP}}(t - vT_{\text{PRI}}) = \text{Re}(x_s(t) e^{j2\pi f_0 t}), \quad t \in \mathbb{R}_{\geq 0}, \quad (4)$$

where $\mathcal{E}_s$ is the transmit power gain and $x_s(t) \in \mathbb{C}$ is the complex-valued lowpass-equivalent transmit signal.

The received bandpass signal $y_s^{\text{BP}}(t) = \text{Re}(y_s(t) e^{j2\pi f_0 t})$ can be defined by the lowpass-equivalent received signal $y_s(t) \in \mathbb{C}$, which can be written as

$$y_s(t) = g(t) * x_s(t) + w_s(t) \quad (6)$$

$$= \sum_{k=1}^K \sum_{\ell=1}^{L_k} \mathbf{g}_{k,\ell} x_s(t - \xi_{k,\ell}(t)) + w_s(t), \quad (7)$$

where $g(t)$ is the complex-valued lowpass-equivalent channel, and $w_s(t) \sim \mathcal{N}(0, \sigma_w^2) \in \mathbb{C}$ is the received noise at the wireless sensing unit. K is the number of targets in the environment, and L_k is the number of channel paths interacting with k^{th} target. $\mathbf{g}_{k,\ell} \in \mathbb{C}$ is the complex-valued channel gain of the ℓ^{th} path of the k^{th} target. $\xi_{k,\ell} = R_{k,\ell}/\varsigma$ is the propagation delay, where ς is the speed of light, and $R_{k,\ell}$ is the propagation distance traveled by the ℓ^{th} path of the k^{th} target. For further details of the channel model, please refer to [4].

At the radar receiver, the received bandpass signal $y_s^{\text{BP}}(t)$ is first passed through a quadrature mixer, and mixed with two versions of the transmit bandpass signal $x_s^{\text{BP}}(t)$, one with a -90° phase shift. Then, low-pass filters and analog-to-digital converters (ADCs) are applied to the outputs of the mixer to generate the in-phase signal $I[u, v]$ and the quadrature-phase signal $Q[u, v]$, for the ADC sample $u \in \mathcal{U}, \mathcal{U} = \{0, 1, \dots, (M_{\text{sample}} - 1)\}$, and for the chirp $v \in \mathcal{V}, \mathcal{V} = \{0, 1, \dots, (M_{\text{chirp}} - 1)\}$. M_{sample} is the number of ADC samples per chirp. Let F_S denote the ADC sampling frequency. The in-phase and quadrature-phase signals are sampled at time $t = uT_S + vT_{\text{PRI}}$, $T_S = 1/F_S$. Finally, the received baseband digital signal $z[u, v] = I[u, v] + jQ[u, v]$ can be formulated as

$$z[u, v] = \sum_{k=1}^K \sum_{\ell=1}^{L_k} \sqrt{\rho_{k,\ell}} e^{-j\vartheta_{k,\ell}} e^{+j\Xi_{k,\ell}} + w_s[u, v] e^{j\chi[u]}, \quad (8)$$

Where $\chi[u] = 2\pi f_0 t_{\text{fast}} + \pi S t_{\text{fast}}^2$ and $t_{\text{fast}} = uT_S$. The channel path received power and phase are $\rho_{k,\ell} = \mathcal{E}_s |\mathbf{g}_{k,\ell}|^2$ and $\vartheta_{k,\ell} = \arg(\mathbf{g}_{k,\ell})$, respectively. The phase term, $\Xi_{k,\ell} = 2\pi \left(f_0 \xi_{k,\ell} + S t_{\text{fast}} \xi_{k,\ell} - \frac{S}{2} \xi_{k,\ell}^2 \right)$, contains the range information of the target.

III. PROBLEM DEFINITION

In this paper, we aim to position the user in the NLoS area and design the RIS interaction vector for communication based on the depth map estimated by the RIS-aided wireless sensing unit. For the sensing purpose, we adopt the design of the RIS sensing codebook in [4], denoted by $\mathcal{F}_s = \{\psi_m^s : m \in \mathcal{M}_s, \mathcal{M}_s = \{0, \dots, M_s - 1\}\}$. The codebook builds a rectangular sensing grid of reflected directions $\mathcal{O} = \{(\theta_m^{\text{az}}, \theta_m^{\text{ze}})_{m=0}^{M_s-1}\}$ with M_h and M_v beams in the horizontal and vertical dimensions, i.e., $M_s = M_v M_h$. Then, the wireless sensing unit can construct a depth map of resolution M_h pixels wide and M_v pixels high. During the sensing process, the wireless sensing unit sweeps over the codebook, where each RIS beam participates in the transmission and reception of a single chirp. For the m^{th} RIS interaction vector, the received baseband digital signal can be expressed as

$$z[u, m] = \underbrace{\sum_{k=1}^K \sum_{\ell=1}^{L_k} \sqrt{\rho_{k,\ell}} e^{-j\vartheta_{k,\ell}} e^{+j\Xi_{k,\ell}}}_{\text{Received signal}} + \underbrace{w_s[u, m] e^{j\chi[u]}}_{\text{Noise}}. \quad (9)$$

By stacking the received M_{sample} ADC samples of each chirp, the received sensing signal matrix $\mathbf{Z} \in \mathbb{C}^{M_{\text{sample}} \times M_s}$ can be constructed as

$$\mathbf{Z} = [\mathbf{z}[0], \mathbf{z}[1], \dots, \mathbf{z}[M_s - 1]], \quad (10)$$

$$\mathbf{z}[m] = [z[0, m], \dots, z[M_{\text{sample}} - 1, m]]^T. \quad (11)$$

To obtain the depth map, we adopt the approach in [4] to process the received signals of the beams and estimate the

depth value of each pixel. The estimated depth map $\mathbf{D}_{\text{map}} \in \mathbb{R}^{M_v \times M_h}$ can be formulated as

$$\mathbf{D}_{\text{map}} = \mathbf{p}_d(\mathbf{Z}; \mathcal{F}_s), \quad (12)$$

where $\mathbf{p}_d(\cdot)$ denote the depth map estimation function.

Next, our objective is to develop a user detection function $\mathbf{p}_u(\cdot)$ that is able to estimate the azimuth and zenith angles towards the user, relative to the RIS, given by

$$(\tilde{\theta}_{\text{UE}}^{\text{az}}, \tilde{\theta}_{\text{UE}}^{\text{ze}}) = \mathbf{p}_u(\mathbf{D}_{\text{map}}). \quad (13)$$

With the acquired angle information of the user, we then propose to design an RIS interaction vector to maximize the received SNR at the UE. Given the communication model described in Section II, the received SNR can be written as

$$\text{SNR} = \frac{\mathcal{E}_c}{\sigma_w^2} |(\mathbf{h}_R \odot \mathbf{h}_T)^T \psi^c|^2. \quad (14)$$

Thus, the optimal RIS interaction vector for communication can be obtained by solving the following optimization problem

$$\begin{aligned} \max_{\psi^c} \quad & |(\mathbf{h}_R \odot \mathbf{h}_T)^T \psi^c| \\ \text{s.t.} \quad & |[\psi^c]_n| = 1, \forall n \in \{1, \dots, N\}. \end{aligned} \quad (15)$$

If we assume that the RIS has a pre-determined communication codebook $\mathcal{F}_c, |\mathcal{F}_c| = M_c$, e.g., a beamsteering/DFT codebook, the objective of the optimization problem becomes finding the optimal beam index in the codebook, which can be formulated as follows:

$$m^* = \arg \max_{\psi_m^c \in \mathcal{F}_c} |(\mathbf{h}_R \odot \mathbf{h}_T)^T \psi_m^c| \quad (16)$$

where the solution can be obtained by an exhaustive search, i.e., beam sweeping, over all the beams in the communication codebook. However, the large codebook would result in a significant beam training overhead. Therefore, we propose to leverage the sensing to reduce the number of trials, which will be presented in the next section.

IV. PROPOSED SOLUTIONS

In this section, we first introduce the proposed approach to positioning the user in the depth map. Next, we describe the RIS interaction vector design for communication. The overall procedure of the proposed solutions are shown in Fig. 2.

A. User Detection with Background Subtraction

The main concept of user detection is to extract useful user information from the depth map, e.g., the user's pixels, which can further be leveraged for communication purposes. To achieve this goal, we design the following key steps.

Background Subtraction: For user detection, the objective is to identify which pixels in the depth map belong to the user. This can be done by subtracting the background, which effectively removes most of the unwanted regions in the depth map. Such background depth map can be, for instance, estimated during the offline stage where there is no user in

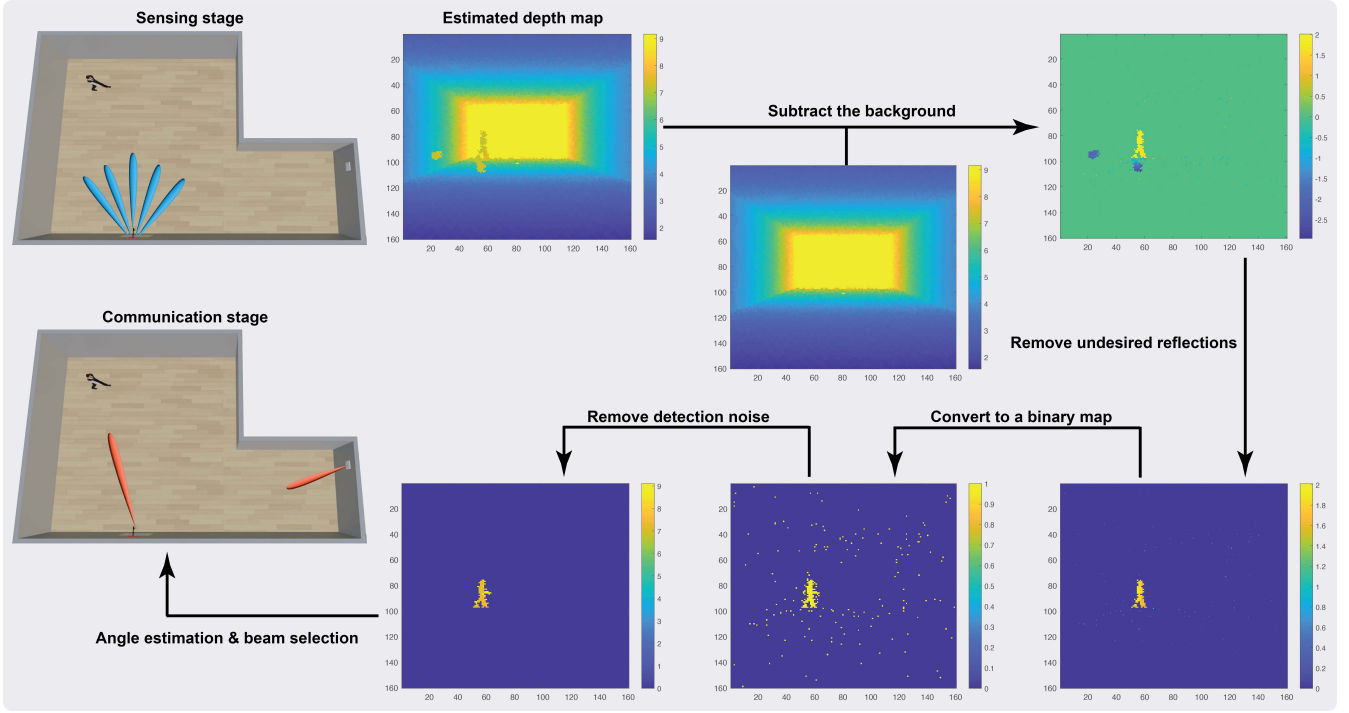


Fig. 2. This figure presents the operation flow of the proposed image-aided communication solution. To begin, the estimated depth map undergoes background subtraction, followed by the elimination of undesired reflections and sensing noise. The pixel coordinates and the corresponding angles of the user can then be obtained from the processed depth map. Finally, the beam selection strategy is performed based on the estimated angles of the user.

the scene. Let $\mathbf{D}_{\text{map}}^b$ denote the background depth map. The background-subtracted depth map $\mathbf{D}_{\text{map}}^{bs}$ can be obtained by

$$\mathbf{D}_{\text{map}}^{bs} = \mathbf{D}_{\text{map}}^b - \mathbf{D}_{\text{map}}. \quad (17)$$

Note that, we subtract the estimated depth map $\mathbf{D}_{\text{map}}$ from the background $\mathbf{D}_{\text{map}}^b$ to make the user's pixels have positive values in the background-subtracted depth map $\mathbf{D}_{\text{map}}^{bs}$.

Removal of Undesired Reflections: Ideally, each beam in the RIS sensing codebook should detect the range/depth of the single-backscattering path of its pointing direction. However, some undesired reflections provide larger channel path gain than the single-backscattering paths, resulting in higher estimated depth values. As a result, there are some regions with negative values in the background-subtracted depth map $\mathbf{D}_{\text{map}}^{bs}$ as shown in Fig. 2. To remove these undesired reflections, we can set the negative values in $\mathbf{D}_{\text{map}}^{bs}$ to zero. This has no impact on the detected user's pixels from the previous step.

Elimination of Detection Noise: Due to the time-varying noise in the sensing process, the background-subtracted depth map suffers from the detection noise, which can be better observed after converting to a binary map. To resolve this, we first identify the pixel coordinates with positive values in the background-subtracted depth map. Next, since the noise pixels are sparse, we leverage a density-based clustering algorithm, DBSCAN [7], to separate the user from the pixels. In the single-user scenario, we choose the cluster with the highest number of elements as the detected user. To determine the azimuth and zenith angles towards the user, we calculate the

rounded mean of the user's pixel coordinates, denoted by (x_u, y_u) , representing the detected user. Given that each pixel is estimated by a beam of a pre-defined reflected direction, we can obtain the angles towards the user using the corresponding pixel coordinate, i.e., $\tilde{\theta}_{\text{UE}}^{\text{az}} = \theta_{m_u}^{\text{az}}$, $\tilde{\theta}_{\text{UE}}^{\text{ze}} = \theta_{m_u}^{\text{ze}}$, $m_u = x_u + M_h y_u$, where $(\theta_{m_u}^{\text{az}}, \theta_{m_u}^{\text{ze}}) \in \mathcal{O}$.

B. RIS Interaction Vector Design for Communication

To design the RIS interaction vector for communication, we can first decompose it into the AP-side and the UE-side beams [8] as shown by

$$\psi^c = \psi_{\text{AP}}^c \odot \psi_{\text{UE}}^c, \quad (18)$$

where ψ_{AP}^c and ψ_{UE}^c denote the AP-side and UE-side RIS interaction vectors. Accordingly, the optimization problem in (15) can be rewritten as

$$\begin{aligned} \max_{\psi_{\text{AP}}^c, \psi_{\text{UE}}^c} & |(\mathbf{h}_T \odot \psi_{\text{AP}}^c)^T (\mathbf{h}_R \odot \psi_{\text{UE}}^c)| \\ \text{s.t.} & \quad |[\psi_{\text{AP}}^c]_n| = 1, \\ & \quad |[\psi_{\text{UE}}^c]_n| = 1, \forall n \in \{1, \dots, N\}. \end{aligned} \quad (19)$$

In mmWave communications, the line-of-sight (LoS) path provides the dominant channel gain. Assuming the locations of the RIS and the AP are known in advance, we can design the AP-side beam to match the LoS propagation path. Thus, the AP-side RIS interaction vector can be expressed as

$$\psi_{\text{AP}}^c = \mathbf{a}^*(\theta_{\text{AP}}^{\text{az}}, \theta_{\text{AP}}^{\text{ze}}), \quad (20)$$

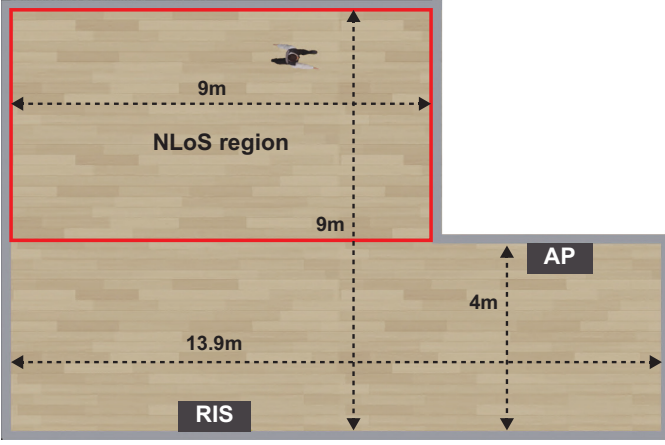


Fig. 3. This figure depicts the adopted indoor scenario. We place the user in the NLoS area (red rectangle) to ensure that there is no direct link between the AP and the UE.

where $\mathbf{a}(\cdot)$ is the far-field RIS array response vector, and $\theta_{\text{AP}}^{\text{az}}$, $\theta_{\text{AP}}^{\text{ze}}$ denote the azimuth and zenith angles towards the AP, relative to the RIS. Similarly, the UE-side RIS interaction vector ψ_{UE}^c can be designed to focus on the LoS path between the RIS and the UE. With the estimated angles towards the user $\tilde{\theta}_{\text{UE}}^{\text{az}}$, $\tilde{\theta}_{\text{UE}}^{\text{ze}}$, the RIS interaction vector can be written as

$$\tilde{\psi}^c = (\mathbf{a}(\theta_{\text{AP}}^{\text{az}}, \theta_{\text{AP}}^{\text{ze}}) \odot \mathbf{a}(\tilde{\theta}_{\text{UE}}^{\text{az}}, \tilde{\theta}_{\text{UE}}^{\text{ze}}))^*. \quad (21)$$

However, this approach requires accurate angle estimation, which may not be feasible in all circumstances due to the sensing noise. Therefore, we then propose an RIS beam selection scheme based on a pre-defined codebook, where a set of candidate beams can be found.

Considering the codebook constraint, we can determine the beam index by calculating the similarities between the RIS interaction vector $\tilde{\psi}^c$ and the beams in the codebook. Then, the selected beam index $\tilde{m}$ is represented as

$$\tilde{m} = \arg \max_{\psi_m^c \in \mathcal{F}_c} |(\tilde{\psi}^c)^H \psi_m^c| \quad (22)$$

Note that, we can find a set of candidate beams by sorting the codebook based on the calculated similarities in (22).

V. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed solutions for the integrated imaging and communication system. We first describe the adopted simulation framework and then present the performance of the RIS beam selection in an indoor scenario.

A. Simulation Framework

In this paper, we propose to leverage depth estimation to detect the user and design an RIS interaction vector for communication. It is essential to use realistic channels in the simulation since the sensing and communication performance highly relies on the environment geometry, scatterers' materials, etc. To that end, we follow the simulation framework in [4] to generate the channels by accurate ray-tracing. Specifically,

we first utilize Blender [9], a high-fidelity 3D graphics design engine, to build a floor plan with a sufficient number of facets. Then, we export the designed floor plan to an accurate 3D ray-tracing simulator, Wireless Insite [10].

For the RIS, we adopt a 40×40 uniform planar array structure at the mmWave 60 GHz operating band. The radar cross section gain of the RIS elements is assumed to be an isotropic gain with half-wavelength RIS element spacing. For the RIS-aided depth estimation, we generate the sensing channel paths and construct the received baseband digital signals (9). We consider an RIS sensing codebook with oversampling factors of four in both vertical and horizontal dimensions, i.e. resolution of 160×160 pixels. The configurations of the FMCW radar follow the settings in [4]. For the RIS-aided communication, we generate the channel between the AP and the RIS, $\mathbf{h}_T$, and the channel between the RIS and the UE, $\mathbf{h}_R$. Then, the composite channel $\mathbf{h}_T \odot \mathbf{h}_R$ can be calculated.

B. Results for An Indoor Scenario

In Fig. 3, we present the top view of the adopted scenario and mark the locations of the AP and the RIS. We consider an L-shape indoor space, where a 1.8m tall person is standing in the NLoS area. The materials of the objects/surfaces, including the concrete walls, floorboard, and ceiling board, are set to the ITU default parameter values at 60 GHz. We generate the samples by placing the person model at 32 distinct locations in the NLoS region, with equal spacing. For each user location, we estimate the depth map and apply the user detection solution as described in Section IV-A. Further, to consider practical scenarios, for each user location, we place the UE's receive antenna at either the jacket pocket or the front/back pants pocket on the person model, resulting in a total of 96 generated samples. We adopt a classical beamsteering codebook to evaluate the performance of the proposed RIS beam selection approach. For comparison, we generate a beamsteering codebook with oversampling factors of four in both elevation and azimuth dimensions.

Fig. 4 shows the top-k normalized beamforming gain of the RIS beam selection. The selection of top-k beams is determined based on the similarities defined in (22). Specifically, we sort the codebook beams based on the similarity values in a descending order to find the top-k beams. The normalized beamforming gain is calculated as the ratio of beamforming gain to the equal-gain beamforming gain. Note that, the equal-gain beamforming is assumed to have perfect channel knowledge. As the value of k increases, better beamforming gain can be achieved with the proposed RIS beam selection approach. In particular, **with the oversampled codebook, the proposed solutions have comparable performance to the optimal beam of the codebook, while requiring less than 0.1% of the beam training overhead of the exhaustive search, which demands 25600 iterations.** Compared to the codebook without oversampling, the oversampled codebook yields a 3 dB gain for the top-25 selected beams, which implies that better received SNR can be attained. This is because the RIS with a large number of elements enables narrow



Fig. 4. This figure presents the beamforming gain provided by the selected top-k beams, compared to the equal-gain beamforming (upperbound) and the exhaustive search. The oversampled codebook is generated with the oversampling factors (OSFs) of four in azimuth and elevation dimensions.

reflected beams, and the non-oversampled codebook can not provide high beamforming gain in all directions. In short, **the simulation results demonstrate the potential of using depth estimation to assist the RIS interaction vector design in the integrated imaging and communication system.**

VI. CONCLUSION

In this paper, we investigate an RIS-aided integrated imaging and communication system that leverages scene depth estimation to achieve low RIS beam training overhead for communication. Specifically, we propose a user detection al-

gorithm to position the user in an estimated depth map, which can be used in the RIS interaction vector design. Then, we design an RIS beam selection scheme based on a pre-defined codebook for communication. Simulation results reveal that the proposed solutions can overcome the large search space of the RIS interaction codebook with high beamforming gain and low overhead. This demonstrates the potential of imaging-aided communication in the proposed ISAC system.

REFERENCES

- [1] U. Demirhan and A. Alkhateeb, "Integrated Sensing and Communication for 6G: Ten Key Machine Learning Roles," *IEEE Communications Magazine*, vol. 61, no. 5, pp. 113–119, 2023.
- [2] Z.-M. Jiang, M. Rihan, P. Zhang, L. Huang *et al.*, "Intelligent Reflecting Surface Aided Dual-Function Radar and Communication System," *IEEE Systems Journal*, vol. 16, no. 1, pp. 475–486, 2022.
- [3] Y. He, Y. Cai, H. Mao, and G. Yu, "RIS-Assisted Communication Radar Coexistence: Joint Beamforming Design and Analysis," *IEEE J. Sel. Areas Commun.*, vol. 40, no. 7, pp. 2131–2145, 2022.
- [4] A. Taha, H. Luo, and A. Alkhateeb, "Reconfigurable Intelligent Surface Aided Wireless Sensing for Scene Depth Estimation," in *IEEE International Conference on Communications*, 2023, pp. 491–497.
- [5] K. Ramasubramanian and T. Instruments, "Using a Complex-Baseband Architecture in FMCW Radar Systems," vol. 19, 2017.
- [6] X. Li, X. Wang, Q. Yang, and S. Fu, "Signal processing for TDM MIMO FMCW millimeter-wave radar sensors," *IEEE Access*, vol. 9, pp. 167 959–167 971, 2021.
- [7] M. Ester, H.-P. Kriegel, J. Sander, X. Xu *et al.*, "A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise," in *Proc. of the 2nd Int. Conf. on Knowledge Discovery and Data Mining*, vol. 96, no. 34, 1996, pp. 226–231.
- [8] S. Jiang, A. Hindy, and A. Alkhateeb, "Sensing Aided Reconfigurable Intelligent Surfaces for 3GPP 5G Transparent Operation," *IEEE Transactions on Communications*, vol. 71, no. 11, pp. 6348–6362, 2023.
- [9] B. O. Community, *Blender - A 3D Modelling and Rendering Package*, Blender Foundation. [Online]. Available: <http://www.blender.org>
- [10] Remcom, "Wireless InSite," <http://www.remcom.com/wireless-insite>.