Millimeter Wave MIMO-Based Depth Maps for Wireless Virtual and Augmented Reality

ABDELRAHMAN TAHA[1], **QI QU**[2], **SAM ALEX**[2], **PING WANG**[2],
WILLIAM L. ABBOTT[2], **AND AHMED ALKHATEEB**[1]

[1]School of Electrical, Computer and Energy Engineering, Arizona State University, Tempe, AZ 85287, USA
[2]Facebook, Inc., Menlo Park, CA 94025, USA

Corresponding author: Ahmed Alkhateeb (aalkhateeb@asu.edu)

ABSTRACT Augmented and virtual reality systems (AR/VR) are rapidly becoming key components of the wireless landscape. For immersive AR/VR experience, these devices should be able to construct accurate depth perception of the surrounding environment. Current AR/VR devices rely heavily on using RGB-D depth cameras to achieve this goal. The performance of these depth cameras, however, has clear limitations in several scenarios, such as the cases with shiny objects, dark surfaces, and abrupt color transition among other limitations. In this paper, we propose a novel solution for AR/VR depth map construction using mmWave MIMO communication transceivers. This is motivated by the deployment of advanced mmWave communication systems in future AR/VR devices for meeting the high data rate demands and by the interesting propagation characteristics of mmWave signals. Accounting for the constraints on these systems, we develop a comprehensive framework for constructing accurate and high-resolution depth maps using mmWave systems. In this framework, we developed new sensing beamforming codebook approaches that are specific for the depth map construction objective. Using these codebooks, and leveraging tools from successive interference cancellation, we develop a joint beam processing approach that can construct high-resolution depth maps using practical mmWave antenna arrays. Extensive simulation results highlight the potential of the proposed solution in building accurate depth maps. Further, these simulations show the promising gains of mmWave based depth perception compared to RGB-based approaches in several important use cases.

INDEX TERMS Beamforming codebooks, depth maps, millimeter wave, virtual reality.

I. INTRODUCTION

Wireless augmented and virtual reality (AR/VR) applications are recently attracting increasing interest. Realizing wireless AR/VR in practice can open the door for a wide range of interesting applications and use cases. Enabling Immersive AR/VR experience, however, requires high resolution and accurate depth perception. This can potentially allow the wireless AR/VR users to move freely within their indoor or outdoor environment. Current depth perception approaches for AR/VR systems rely mainly on RGB-D (depth) cameras for constructing the depth maps. While RGB-D based depth map construction approaches can generally provide good accuracy, they suffer from critical limitations in scenarios

with bright shiny or transparent surfaces, dark objects, and large rooms among others. These limitations stem from the fundamental properties of the way visible light propagate and interact with the different surfaces.

In order to overcome these limitations, **we propose to leverage mmWave systems and signals for improving the depth map estimation accuracy**. This is motivated by the interesting characteristics of mmWave signals and by the note that mmWave systems will be deployed in future AR/VR devices anyway for meeting the wireless communication requirements [1]. In terms of the mmWave signal characteristics, the propagation of these signals is not affected by the interference from the light sources which makes mmWave systems capable of detecting bright and dark objects. Further, the mmWave diffuse scattering and specular reflection properties could help in detecting transparent objects as well as

The associate editor coordinating the review of this manuscript and approving it for publication was Yunlong Cai.

rough surfaces. These aspects among others motivate exploring the potential of leveraging mmWave transceivers for complementing the RGB-D depth-maps in AR/VR systems, which is the focus of this paper.

A. PRIOR WORK

Previous depth map construction approaches focused on leveraging: (i) monocular images using RGB cameras [2], (ii) passive/active stereo images using either RGB-D depth cameras [3], [4] or infrared (IR) stereo cameras [5], [6], and (iii) gated images using active gated imaging cameras [7], [8]. In [2], a monocular depth estimation approach capable of capturing the object boundaries is proposed. In [3], RGB images along with sparse depth samples, acquired from depth cameras or computed via Simultaneous Localization and Mapping (SLAM) algorithms, are used jointly to reconstruct the depth maps. An alternative approach for depth estimation was proposed in [4], where a monocular structured-light camera — a calibrated stereo set-up with one camera and one laser projector— is leveraged for estimating the disparity. As for the active stereo systems, in [5], IR projected pattern from stereo IR cameras is adopted for depth estimation through active stereo matching. The IR images are acquired from the Intel Realsense camera [9]. Also, the IR pattern characteristics needed for active stereo matching are described in [6]. In addition, high-resolution depth images can be achieved for far objects using active gated imaging systems, as in [7], [8].

These depth map construction approaches [2]–[8], [10], however, have several important limitations complications as follows. (i) First, these depth map construction approaches normally fail to sense the depth for shiny, dark, transparent, and distant surfaces. While there are some attempts in solving these challenges using IR stereo cameras [5] or excessive processing of the RGB-D images [11], there is no complete and general solution yet to this problem. (ii) Further, these IR and RGB-D based depth map construction algorithms suffer from a critical limitation, which is the depth ambiguity for far objects/surfaces. The depths for distant surfaces can not be resolved by the algorithms in [5], [11]. (iii) Another key challenge is the additional bill of materials (BOM) cost incurred from integrating the IR stereo camera systems into the wireless AR/VR device architectures. On the contrary, the existing mmWave systems in the wireless AR/VR device architectures incurs no additional BOM cost when leveraged for depth map estimation purposes jointly with the primary purpose of wireless communications. (iv) The field of view coverage is also a main challenge. The depth map coverage is limited by the camera field of view. The camera field of view is constrained by the camera lens and by the light sensor. The field of view in mmWave MIMO systems, however, is constrained by the array radiation pattern, as will be explained in Section VI. By contrast, the typical field of view in mmWave MIMO systems can be larger than the typical camera field of view.

These challenges motivate the research for other technologies to complement the RGB-D cameras in accurately sensing the VR/AR environment. One promising technology for this goal is employing wireless millimeter wave (mmWave) systems. Since mmWave antenna arrays will be used to satisfy the communication high data rate demands of wireless VR/AR, it is interesting to investigate if they could also be useful for VR/AR-relevant sensing functions, such as depth estimation. Initial studies for using mmWave communication arrays for radar and sensing were presented in [12], [13]. These studies, however, focused only on the ranging problem (of one or multiple targets), not on the depth map construction problem. Other mmWave sensing and tracking work that was not restricted to communications hardware was presented in [14], [15]. The research in [14], [15], though, targeted tracking a single object in a small distance, and cannot be directly applied to depth estimation of surrounding surfaces in VR/AR. Further, the work in [12]–[15], did not study the trade-offs between estimation accuracy and different system parameters, such as number of antennas and adopted bandwidth, and did not compare between the system performance under transceiver architectures constraints, such as those imposed on the analog phased-array transceiver architectures. By contrast, interesting research challenges are accompanying the mmWave MIMO based scene depth map construction framework ranging from beam codebook design challenges to scene depth estimation challenges. These challenges will be addressed in this work and will be explained in detail in Section V.

B. CONTRIBUTION

In this paper, we consider the mmWave MIMO based depth map construction problem for AR/VR systems, adopting mmWave communication hardware and frame structure. The contributions of this paper can be summarized as follows.

- *mmWave MIMO depth map construction framework:* We formulate the mmWave MIMO depth map construction problem and propose a general framework for building depth maps under the constraints imposed by mmWave communication hardware and frame structure.
- *A design for depth-map suitable sensing beamforming codebook:* We define the characteristics of the desirable mmWave sensing beamforming codebook for efficient depth map construction and develop a codebook construction approach that meets these characteristics.
- *High-resolution depth map construction approach:* Given the designed beamforming codebook, we develop a novel signal processing approach for jointly processing the signals received by the sensing beams and building high-resolution depth maps.

The proposed solution is extensively evaluated using accurate ray-tracing channels generated from Wireless InSite [16], and ground truth depth images generated from Blender [17]. The simulation results show the promise of mmWave MIMO sensing in becoming a viable depth estimation solution for communication-constrained sensing systems, either as a standalone approach or as an integrated approach with RGB-D
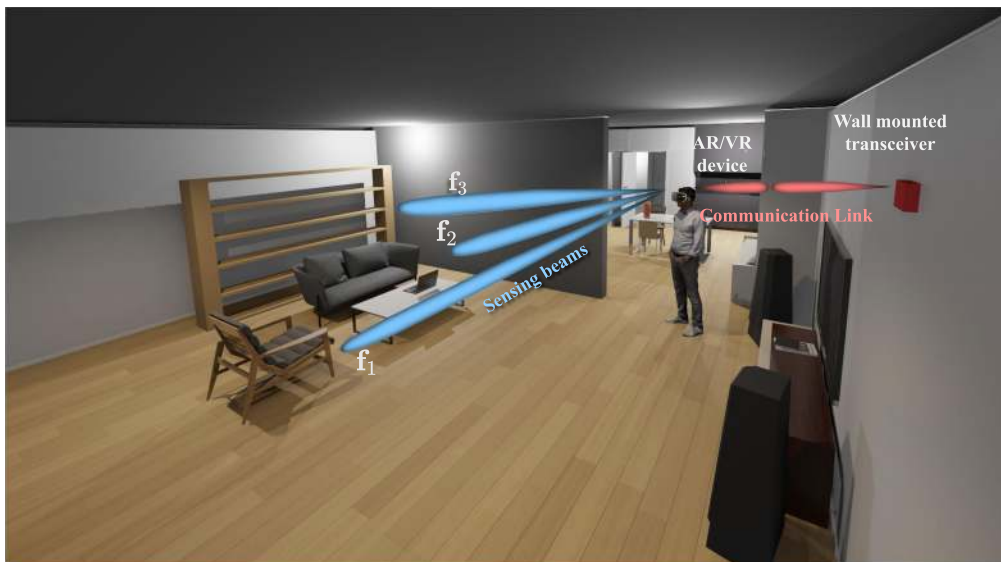
FIGURE 1. The considered setup where the mmWave communication system, deployed at the AR/VR device, is jointly leveraged for sensing and depth map construction. This figure is generated using Blender [17] with 3D models downloaded from [18]–[21].

depth cameras. These simulation results can be of great usefulness for various applications; they can be generally applied to AR/VR devices, smart home devices, or auto drive devices.

Notation: We use the following notation throughout this paper: $\mathbf{A}$ is a matrix, $\mathbf{a}$ is a vector, a is a scalar, $\mathcal{A}$ is a set of scalars, and $\boldsymbol{\mathcal{A}}$ is a set of vectors/matrices. $\|\mathbf{a}\|_p$ is the p-norm of $\mathbf{a}$. $|\mathbf{A}|$ is the determinant of $\mathbf{A}$, $\|\mathbf{A}\|_F$ is its Frobenius norm, whereas $\mathbf{A}^T$, $\mathbf{A}^H$, $\mathbf{A}^*$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$ are its transpose, Hermitian (conjugate transpose), conjugate, inverse, and pseudo-inverse respectively. $[\mathbf{A}]_{r,:}$ and $[\mathbf{A}]_{:,c}$ are the rth row and c^{th} column of the matrix $\mathbf{A}$, respectively. diag($\mathbf{a}$) is a diagonal matrix with the entries of $\mathbf{a}$ on its diagonal. $\mathbf{I}$ is the identity matrix. $\mathbf{1}_N$ and $\mathbf{0}_N$ are the N-dimensional all-ones and all-zeros vector respectively. $\mathbf{A} \otimes \mathbf{B}$ is the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, $\mathbf{A} \circ \mathbf{B}$ is their Khatri-Rao product, and $\mathbf{A} \odot \mathbf{B}$ is their Hadamard product. $\mathcal{N}(\mathbf{m}, \mathbf{R})$ is a complex Gaussian random vector with mean $\mathbf{m}$ and covariance $\mathbf{R}$. $\mathbb{E}[\cdot]$ is used to denote expectation. vec($\mathbf{A}$) is a vector whose elements are the stacked columns of matrix $\mathbf{A}$.

II. SYSTEM AND CHANNEL MODELS
In this section, the system model for the adopted communication-constrained sensing framework is first formulated, followed by the characterization of the adopted channel model.

A. SYSTEM MODEL
In this paper, we propose to reuse the same AR/VR mmWave communication system/circuits to do the sensing and depth map construction, as shown in Fig. 1. Hence, we adopt a sensing model that accounts for the mmWave communication system/circuit constraints. This communication-constrained sensing model consists of a transmitter and a receiver; both are connected through a self-isolation circuitry to a shared N antenna array, as depicted in Fig. 2. This type of operation is commonly referred to as MIMO in-band

full-duplex operation [22]. We assume that the transmitter and receiver chains are well-isolated by an isolation circuitry to avoid any self-interference. This assumption is reasonable with the recent developments of self-interference systems. One example of these systems is the magnetic-free non-reciprocal circulators (i) based on coupled-resonator loops [23] or (ii) based on CMOS circulators operating in the 28GHz mmWave band [24]. Another example is the receiver with integrated magnetic-free non-reciprocal circulator and baseband self-interference cancellation operating in the Sub-6 GHz band [25]. A third example is the magnetic-free SOI CMOS circulator operating in the 60GHz mmWave band [26]. Accounting for this self-interference, however, is an important direction for future extensions.

Further, and for the sake of having low-cost and power consumption mmWave transceivers, we adopt an analog-only architecture for the N-antenna array used for transmission and reception, [27], [28], where the beamforming/combining is done in the analog domain using a network of phase shifters. Next, we summarize the transmit and receive signal models.

1) TRANSMIT SIGNAL MODEL
We consider a wideband single-carrier waveform comprising multiple time frames. These frames are transmitted over an aggregated time interval of T seconds during which the environment is assumed to be relatively static. This time interval is commonly referred to as a coherent processing interval (CPI) [29]. Each frame consists of both data and preamble sequences designed for the wireless communication function. The co-existing sensing model also uses these preamble sequences to sense the environment and build the depth maps, as will be explained in detail in the following sections. This can be achieved by either splitting the frames between sensing and communication or by designing the
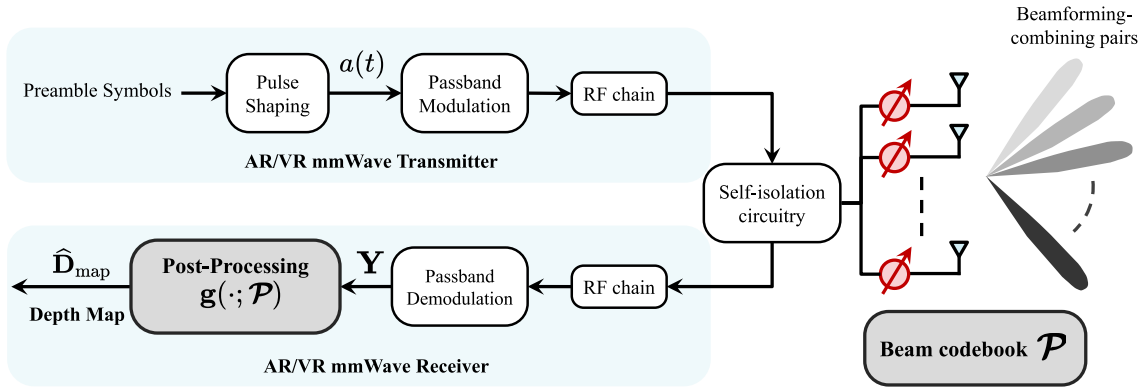
FIGURE 2. A block diagram of the communication-constrained sensing model is illustrated. The sensing framework, Π, consists of (a) the beam codebook design $\mathcal{P}$ and (b) the post-processing design $\mathbf{g}(., \mathcal{P})$, to estimate the scene depth map $\hat{\mathcal{D}}$. The upper path represents the transmitter path, while the lower path represents the receiver path.

sensing and communication beam training operations to share the same preamble sequences. Next, for ease of exposition, we assume that M frames/preamble sequences are dedicated for sensing. If $s_m[n]$ denotes the nth transmitted symbol at the mth frame, with $\mathbb{E}[|s_m[n]|^2] = 1$, then the complex-baseband representation of the transmit waveform can be written as [30]

$$a(t) = \sqrt{\mathcal{E}_\text{s}} \sum_{m=0}^{M-1} \sum_{n=0}^{N_m-1} s_m[n]\, \delta(t - nT_\text{S} - mT_\text{F}), \quad (1)$$

where $\mathcal{E}_\text{s}$ represents the average energy per symbol, T_S is the symbol time, and T_F is the frame duration. N_m is the number of symbols in the mth frame, which is divided into a preamble sequence of length N^p and a set of data symbols of length N_m^d. Further, we assume that the same preamble sequence $s^\text{p}[n], n \in \{1, \ldots, N^\text{p}\}$, is transmitted in the first N^p symbols of each frame. Note that for the sake of simplifying the transmit and receive signal representation, we incorporated the transmit pulse shaping and receive filtering functions into the channel model. Finally, if a beamforming vector $\mathbf{f} \in \mathbb{C}^{N \times 1}$ is used to transmit the signal at the AR/VR device, the complex-baseband representation of the transmitted signal can expressed as

$$\mathbf{x}(t) = \mathbf{f}a(t). \quad (2)$$

This transmitted signal will interact with the environment (through reflection, scattering, etc.) and will be received back by the AR/VR device. Next, we describe the receive signal model.

2) RECEIVE SIGNAL MODEL
Let G_tar denote the number of targets/scatterers in the environment. Then, focusing on the preamble sequence transmission/reception (i.e., the first N^p symbols of each frame), the receive *sensing* signal of the mth frame can be written as

$$y_m[n] = \sum_{g=1}^{\mathsf{G}_\text{tar}} \sum_{d=0}^{L_\text{d}-1} \sqrt{\mathcal{E}_\text{s}} \mathbf{w}^H \mathbf{H}_{d,g} \mathbf{f} s^\text{p}[n-d] + \mathbf{w}^H \mathbf{v}_m[n], \quad (3)$$

where $\mathbf{w} \in \mathbb{C}^{N \times 1}$ is the combining vector at the AR/VR, and $\mathbf{v}_m[n] \sim \mathcal{N}_\mathbb{C}(\mathbf{0}, \sigma_\text{n}^2 \mathbf{I})$ is the receive noise with variance σ_n^2. $\mathbf{H}_{d,g} \in \mathbb{C}^{N \times N}, d \in \{1, \ldots, L_\text{d} - 1\}$, is the delay-$d$ channel matrix between the transmission from and the reception by the AR/VR antenna array, which is described in the following subsection.

B. CHANNEL MODEL
Given that the depth sensing problem highly relies on the accurate modeling of the surrounding environment and its geometry, we adopt a geometric channel model in this work. More specifically, we consider the extended Saleh-Valenzuela wideband geometric channel model [31]–[34]. Based on that, the gth target contribution in the delay-d channel, $\mathbf{H}_{d,g}$, can be modeled as

$$\mathbf{H}_{d,g} = \sqrt{G_g} \sum_{\ell=1}^{L_\text{ray}} \Big[\alpha_\ell e^{-j2\pi f_\text{c} \tau_\ell} p(dT_\text{s} - \tau_\ell)$$
$$\times \mathbf{a}_\text{R}\left(\phi_{\ell,g}^\text{R}, \theta_{\ell,g}^\text{R}\right) \mathbf{a}_\text{T}^H\left(\phi_{\ell,g}^\text{T}, \theta_{\ell,g}^\text{T}\right) \Big], \quad (4)$$

where L_ray is the number of channel clusters; each cluster is contributing with one ray of complex channel coefficient α_ℓ, time delay τ_ℓ, and azimuth/elevation angles of departure and arrival, $\phi_{\ell,g}^\text{T}, \theta_{\ell,g}^\text{T}$ and $\phi_{\ell,g}^\text{R}, \theta_{\ell,g}^\text{R}$, respectively. $\mathbf{a}_\text{T}(\cdot, \cdot)$ and $\mathbf{a}_\text{R}(\cdot, \cdot)$ represent the transmit and receive array response vectors associated with the angles of departure and arrival respectively. The transmit and receive pulse shaping signals are included within $p(t)$ such that $p(t) = p_\text{T}(t) * p_\text{R}(t)$. The path gain associated with the gth target is denoted by G_g and can be expressed as

$$G_g = \frac{G_\text{T} G_\text{R} \lambda^2 \sigma_g^\text{RCS}}{(4\pi)^3 (\rho_g)^{2\text{PL}}}, \quad (5)$$

where G_T and G_R are the transmitter and receiver gains, λ is the operating wavelength, PL is the path loss exponent. Finally, ρ_g denotes the distance (range) between the AR/VR device and the gth target/scatterer and σ_g^RCS denotes the radar cross section of this target.

III. PROBLEM DEFINITION

Our objective in this paper is to efficiently estimate the depth/range map of the surrounding environment using the communication-constrained mmWave MIMO sensing model in Section II. Before delving into the formal problem definition, it is important to distinguish between the range and the depth of a certain target. As depicted in Fig. 3, the range of a target with respect to the AR/VR camera (which is aligned with the AR/VR antenna array) is the linear radial distance from the camera center (focal point) to the target. For the depth, it is measured by the *y*-coordinate of the camera center (focal point) with respect to the *x*-*z* plane of the target. Given that the range and depth can be calculated from one another, we focus our formulation on the depth estimation problem. Next, we define the depth map of the surrounding environment with respect to an AR/VR device.
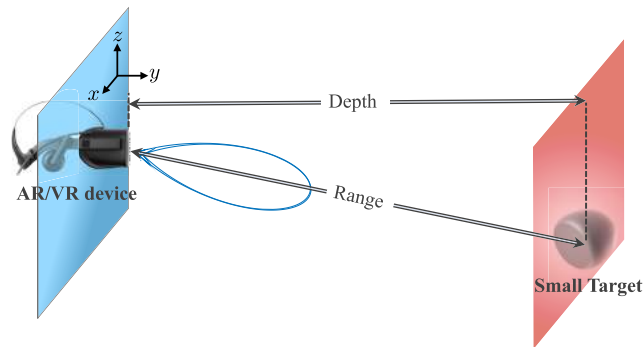
FIGURE 3. This figure shows the conventional single target range estimation problem, where one target exists in free space in line-of-sight (LoS) with the AR/VR device. This device steers perfectly one beam towards that target to estimate the range.

Definition (Depth Map): We define the depth map $\mathbf{D}_{\mathrm{map}}$ *of resolution* $M_h \times M_w$ *as an image of* M_h *pixels high and* M_w *pixels wide, where the value of each pixel represents the smallest depth between the AR/VR device and the targets/objects in this pixel.*

In this paper, we express this depth map as an $M_h \times M_w$ matrix $\mathbf{D}_{\mathrm{map}} \in \mathbb{R}^{M_h \times M_w}$. Further, we use $M_{\mathrm{res}} = M_h M_w$ to denote the total number of pixels in the depth map. The range map $\mathbf{R}_{\mathrm{map}} \in \mathbb{R}^{M_h \times M_w}$ is similarly defined. Now, given the system and channel models in Section II, the AR/VR device constructs the estimated mmWave-based depth map through two main steps: (i) sensing the environment using several beamforming and combining sensing vectors and (ii) post-processing the receive sensing signal to construct the estimated depth map. More formally, if a beamforming-combining pair $(\mathbf{f}_m, \mathbf{w}_m)$ is used to transmit and receive the N^{p} symbols of the mth preamble sequence, then the receive sensing signal can be expressed as

$$y_m[n] = \sum_{d=0}^{L_{\mathsf{d}}-1} \sqrt{\mathcal{E}_{\mathsf{s}}} \mathbf{w}_m^H \mathbf{H}_d \mathbf{f}_m s^{\mathrm{p}}[n-d] + \mathbf{w}^H \mathbf{v}_m[n], \quad (6)$$

where $n \in \{0, 1, \ldots, N^{\mathrm{p}} + L_{\mathsf{d}} - 1\}$, $\mathbf{H}_d = \sum_{g=1}^{\mathsf{G}_{\mathrm{tar}}} \mathbf{H}_{d,g}$. By stacking the $N^{\mathrm{p}} + L_{\mathsf{d}}$ receive symbols (from transmitting the preamble sequence), we get $\mathbf{y}_m = [y_m[0], \ldots, y_m[N^{\mathrm{p}} + L_{\mathsf{d}} - 1]]^T$, which represents the receive sensing vector of one preamble sequence using one beamforming-combining pair. Next, if M preamble sequences are used to sense the environment via M beamforming-combining pairs $(\mathbf{f}_m, \mathbf{w}_m)$, $m \in \{1, \ldots, M\}$, then the aggregated receive sensing signal can be written as

$$\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_M]. \quad (7)$$

For ease of exposition, we define the sensing beamforming codebook $\mathcal{P}$ as the codebook that includes the M beamforming-combining pairs, i.e., $\mathcal{P} = \{(\mathbf{f}_m, \mathbf{w}_m) : m \in \{1, \ldots, M\}\}$. Finally, given the receive sensing matrix $\mathbf{Y}$, a post-processing is applied for estimating the depth map. If $\mathbf{g}(.)$ denotes the post-processing function, the estimated depth map $\widehat{\mathbf{D}}_{\mathrm{map}} \in \mathbb{R}^{M_h \times M_w}$ can be written as

$$\widehat{\mathbf{D}}_{\mathrm{map}} = \mathbf{g}(\mathbf{Y}; \mathcal{P}). \quad (8)$$

Our objective in this paper then is to design the sensing beamforming codebook $\mathcal{P}$ and the post-processing $\mathbf{g}(\cdot)$ to efficiently estimate the depth map $\widehat{\mathbf{D}}_{\mathrm{map}}$ to be as close as possible to the actual depth map $\mathbf{D}_{\mathrm{map}}$. To evaluate the performance of the proposed approaches, we will adopt the root-mean squared estimation error (RMSE) and the mean absolute error (MAE) between the depth maps, which are defined as

$$\Delta_{\mathrm{RMSE}} = \sqrt{\frac{1}{M} \left\| \mathbf{D}_{\mathrm{map}} - \mathbf{g}\left(\mathbf{Y}; \mathcal{P}\right) \right\|_2^2}, \quad (9)$$

$$\Delta_{\mathrm{MAE}} = \frac{1}{M} \left\| \mathbf{D}_{\mathrm{map}} - \mathbf{g}\left(\mathbf{Y}; \mathcal{P}\right) \right\|_1^2. \quad (10)$$

In Section V, we will present the general framework of our proposed depth map estimation approach. This will be followed by a detailed description of the two main components in this framework, namely the beamforming codebook design $\mathcal{P}$ and the post-processing solution $\mathbf{g}(.)$, in Sections VI and VII.

IV. BACKGROUND

Before going into the proposed framework for estimating depth/range maps using mmWave MIMO, we provide a brief background on the basis of the single-target range estimation problem. For a preliminary model, consider one target in the free space with a Line-of-sight (LoS) path to the AR/VR device. Further, consider the case when one mmWave beam is perfectly steered towards that target, as depicted in Fig. 3. Adopting this preliminary model, the target range estimation accuracy bound will be first examined. Then, a description of the main algorithms used in the literature to approach this problem is provided.

A. TARGET RANGE ESTIMATION ACCURACY

Our main objective is to find the fundamental limit for mmWave MIMO based depth estimation, which can be
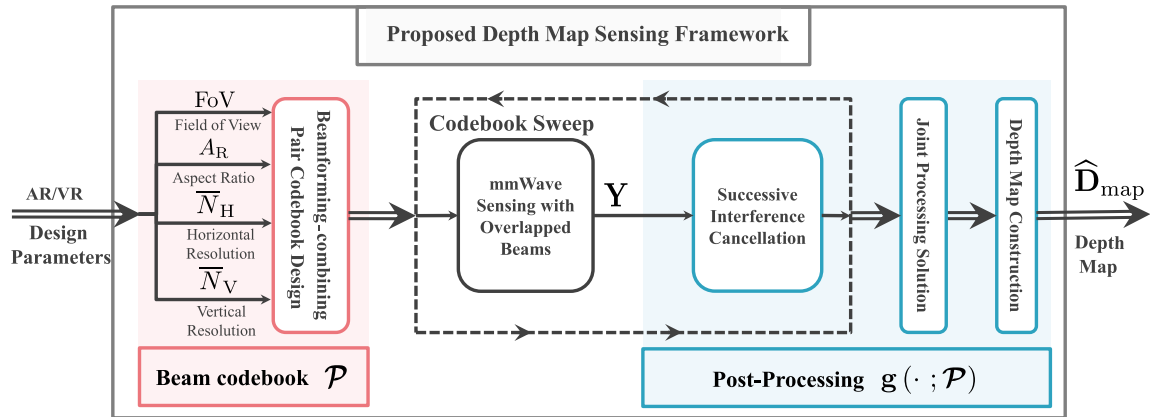
FIGURE 4. The figure summarizes the proposed sensing framework for mmWave MIMO based depth estimation, which involves sensing the scene using the designed beamforming codebook $\mathcal{P}$ and applying the proposed post-processing operations g(.; $\mathcal{P}$) to the receive signal to construct the estimated depth map $\widehat{\mathbf{D}}_{\mathbf{map}}$.

considered as range estimation at every possible eyesight direction, i.e. at every azimuth angle $\phi \in [0, 2\pi[$ and every elevation angle $\theta \in [0, \pi[$. For the range estimation accuracy, one useful metric is the Cramer-Rao lower bound (CRLB) on the range estimation. For white Gaussian noise, the CRLB provides a lower bound on the mean-squared-error of any unbiased estimator, hence it is used as a benchmark for the performance analysis of parameter estimation [35]. Considering the case of range estimation for a single target, the CRLB of this single target range is formulated as [13], [29], [35]

$$\sigma_{\hat{\rho}}^2 \geq \frac{\varsigma^2}{8 P_{\text{int}} \eta^2 B^2 \text{SNR}_{\text{rad}}}, \quad (11)$$

where ς is the speed of light, B is the transmission bandwidth, P_{int} is the integration gain and is equal to the number of symbols used for preamble estimation, and η depends on the power spectral density shape of $a(t)$ over the preamble duration. Under the assumption of a flat spectral density for $a(t)$, $\eta^2 = (2\pi)^2 /12$. The radar signal-to-noise ratio for this target can then be expressed as $\text{SNR}_{\text{rad}} = \mathcal{E}_s G_{\text{rad}}/\sigma_n^2$, where G_{rad} denotes the path gain associated with the target.

B. TARGET RANGE ESTIMATION ALGORITHMS
Estimating the round trip delay $\hat{\tau}$ is equivalent to finding the range estimate $\hat{\rho}$, since they are directly related through $\hat{\tau} = 2\hat{\rho}/\varsigma$. Given the extensive research on delay estimators in the literature [29], [36], we will restrict the scope of this paper on the magnitude based delay estimators in [37] for simplicity. In a general sense, given a known transmit preamble sequence $x_0[n]$ and the received baseband sequence $z[n]$, the receiver can estimate the round-trip delay by maximizing an objective function, the cross-correlation function between the two time-sequences, over a range of possible delays. Based on this notion, two delay estimators are formulated as follows [37].

1) BASIC CORRELATOR
The basic correlator is a coarse delay estimator that performs the maximization at the same sampling frequency, f_S, tuned by the AR/VR communication system. Assume that the length of the received baseband sequence, $z[n]$, is L_z samples, where the last N_z samples are non-zeros. The range estimate can then be formulated as
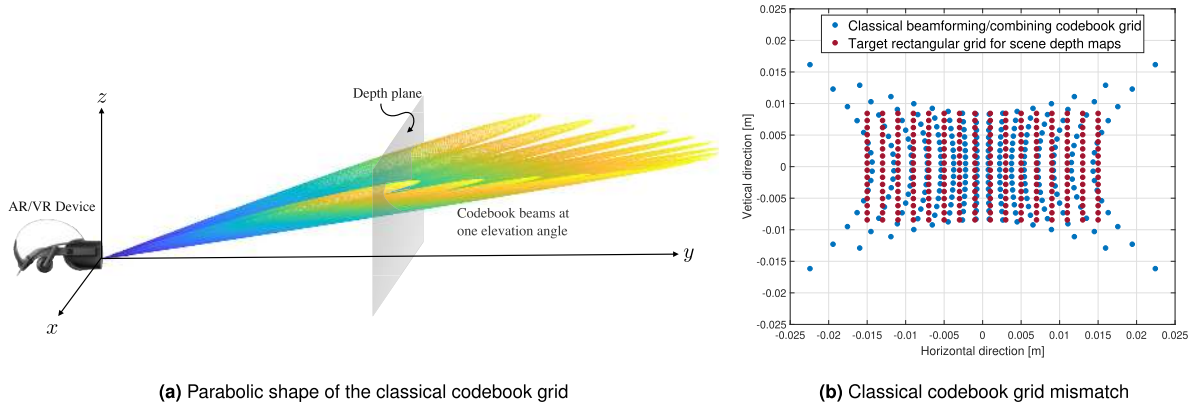
$$\hat{\rho}^{\text{BC}} = \frac{\varsigma T_S}{2} \arg\max_{q:q\in\mathcal{Q}} \left| \sum_{n=L_z-N_z}^{L_z-1} x_0[n] \times z^*[n-q] \right|^2. \quad (12)$$

where $T_S = \frac{1}{f_S} = \frac{1}{B}$ denotes the sampling time, $\mathcal{Q}$ represents the set of possible discrete sample delays, and the optimal q solution is denoted by q^{BC}. Unfortunately, the accuracy of this range estimate is limited by the sampling frequency f_S. One attempt of improving the estimation accuracy is by performing the maximization at a higher sampling frequency. This attempt, however, increases the computational complexity dramatically, which motivates the role of the upcoming delay estimator, the massive correlator [37].

2) MASSIVE CORRELATOR
The primary function of the massive correlator is to perform the maximization of the objective function at a higher sampling frequency without the computational burden of computing the shift in real time. For this reason, [37], [38] introduced the solution of pre-designing a specific correlator bank that contains shifted versions of the reference sequence, $x_0[n]$. The receiver will then multiply the received sequence by the correlator bank to compute the objective function.

We describe the steps of the massive correlator algorithm as follows [37], [38]. (a) Upsample $x_0[n]$ with a sampling frequency higher than f_S, denoted as f_{est}. (b) Define the correlator bank matrix, $\mathbf{X}_0$, where each row of this matrix is a shifted version of the upsampled $x_0[n]$. Let the number of rows in $\mathbf{X}_0$ be equal to $(2\delta + 1)$, where δ is the largest

(a) Parabolic shape of the classical codebook grid

(b) Classical codebook grid mismatch

FIGURE 5. (a) The intersections between the classical codebook beam directions and the *x-z* depth plane form the parabolic shape of the classical codebook grid. (b) The mismatch between the classical codebook grid of a 16 × 16 UPA and the desirable rectangular grid for a depth map is illustrated at a *y* = 13.32mm depth plane, for a scene of 100° field of view and 16/9 aspect ratio.

lag/advance discrete fractional delay in the receive sequence, such that $\delta = \frac{f_{est}}{2f_S}$. (c) Downsample independently each row of the correlator matrix to the lower sampling frequency f_S; let the resulting matrix be named as $\mathbf{B}_0$. The reason for this step is to test delays at the higher sampling frequency, f_{est}, but only apply multiplications at the lower sampling frequency, f_S. (d) Shift back the receive sequence, $z[n]$, with the coarse discrete delay estimate, q^{BC}, such that $\bar{z}[n] = z[n + q^{BC}]$, and then concatenate the sequence into one row vector, $\bar{\mathbf{z}}$. (e) Calculate the fractional range estimate, $\hat{\rho}'$, such that $\hat{\rho}' = \frac{\varsigma}{2f_{est}}\left(-(\delta + 1) + \arg\max_{q'}[\mathbf{g}]_{q'}\right)$, where $\mathbf{g} = \bar{\mathbf{z}} \times \mathbf{B}_0^H$. (f) Calculate the fine range estimate, $\hat{\rho}^{MC}$, such that $\hat{\rho}^{MC} = \hat{\rho}^{BC} + \hat{\rho}'$.

With an end goal of constructing depth maps, these range estimation algorithms will then be leveraged by the mmWave MIMO based depth estimation as explained in the upcoming sections. In the next section, we formulate a general framework for scene depth estimation.

V. GENERAL FRAMEWORK FOR SCENE DEPTH ESTIMATION

In this section, we highlight the key elements of the proposed depth map estimation approach, namely the sensing beamforming codebook $\mathcal{P}$ and post-processing $\mathbf{g}(.)$, and discuss the challenges associated with designing these elements. As depicted in Fig. 4, we first design the sensing beamforming codebook $\mathcal{P}$ offline based on the desired AR/VR properties such as the field of view, the scene aspect ratio, and the number of horizontal and vertical beams covering the scene view. To build the depth map of a certain scene, the beam pairs of the designed codebook are used to sense the environment and acquire the receive sensing matrix $\mathbf{Y}$ in (6). This receive signals are then jointly processed using the post-proposed approach to build the depth map. In the remaining of this section, we explain the challenges associated with designing the codebook and the post-processing operations. Then, we will present how our proposed solutions overcome these challenges in Sections VI and VII.

A. CODEBOOK DESIGN CHALLENGES

To effectively sense the surrounding environment and build efficient depth maps, the beams of the sensing codebooks should be designed to scan the full scene. Since the mmWave MIMO based depth maps may potentially complement the RGB-D based maps, our objective is to build a beamforming codebook that scans the full rectangular grid of the typical depth sensors of the AR/VR cameras. However, the classical beam steering codebooks such as the DFT codebooks [39], that independently sample the azimuth and elevation directions, do not normally fit a rectangular grid. They instead form a parabolic grid, i.e., for a fixed elevation angle, the grid line of these codebook beams are parabolic curves as shown in Fig. 5(a). This mismatch between the mmWave MIMO-based and camera-based depth grids could lead to clear distortion in the joint mmWave/RGB-D depth map construction and make it hard to complement the RGB-D depth map using mmWave MIMO sensing.

One possible solution is to estimate the depths on the parabolic grid using the classical beamforming codebook and then interpolate/extrapolate to calculate the rectangular depth map. The main disadvantage of this solution, however, is that the interpolation can potentially lead to considerable loss in the depth map accuracy as the changes of the depth are not normally smooth in nature. Hence, in order to avoid the interpolation loss, the more persuasive solution is to develop a depth map compatible beamforming codebook that fits exactly the desirable rectangular sensor grid. With this motivation, we propose a beamforming/combining design approach in Section VI to overcome the codebook mismatch challenge.

B. SCENE DEPTH ESTIMATION CHALLENGES

The sensing beamforming codebook is used to sense the surrounding environment. Now, given the receive sensing matrix $\mathbf{Y}$, the objective of the post-processing is to construct an accurate depth map of the facing scene. This process, however, has several challenges. In order to explain these

challenges, let's first consider the case when the environment has only a single target. In this case, the sensing/scanning beam that is directed towards the region that includes this target will result in some backscattering signal. This signal can be used for calculating the round-trip time of flight and consequently the range of this target, leveraging the MIMO radar concepts [29], [40] and the algorithms detailed in Section IV-B. In terms of the range/depth map, the pixel that includes the region of this target will simple have the value of the estimated range/depth. In practice, however, the environment has several targets/surfaces and the mmWave arrays have strict constraints on their hardware: power budget, computational complexity, etc. These limitations lead to critical challenges for our objective of building accurate depth and range maps of the environment. More specifically, if we adopt the approach that scans the surrounding scene using a beamforming codebook and processed the receive sensing signal of each beam independently to estimate the depth of the region defined by this beam, then this approach will have the following key drawbacks.

- **Low-resolution depth-maps:** The low resolution drawback is mainly due to (i) the limitation on the number of AR/VR antennas, which is controlled by many factors in the AR/VR device such as the device dimensions, computational complexity, circuit routing, power consumption, etc., and (ii) the number of beams in the sensing codebook $\mathcal{P}$, which is limited by the time allocated for the depth estimation process.

- **Inter-target interference:** The constraints on the number of antennas at the AR/VR device limit the system spatial resolution. This makes it hard to differentiate between the ranges/depths of the different targets/surfaces that are close to each other. In other words, when measuring the depth of the object in a particular region/ direction, multiple objects/surfaces may reflect the incident signal at the same time. The interference between these reflected/scattered signals may highly affect the accuracy of the range/ depth estimation. Hence, if a certain pixel has multiple objects/surfaces, it will be difficult to estimate the shortest depth of the objects in this pixel (to follow the depth map definition in Section III).

- **Inter-path interference:** When sensing the range/depth of a certain target, the optimal situation (in terms of depth estimation accuracy) is when the target backscatters a single ray to the receiving array. In practice, however, the signal incident on a certain target may experience more than one phenomenon, such as scattering, reflection, diffraction, etc., which results in multiple rays. More than one of these rays could traverse the environment in different ways/directions, especially in indoor environments, before reaching the receiver. This means that they may reach the receiving array from multiple angles and with different time of flights. This causes an inter-path interference which makes it hard to accurately estimate the range/depth of the target
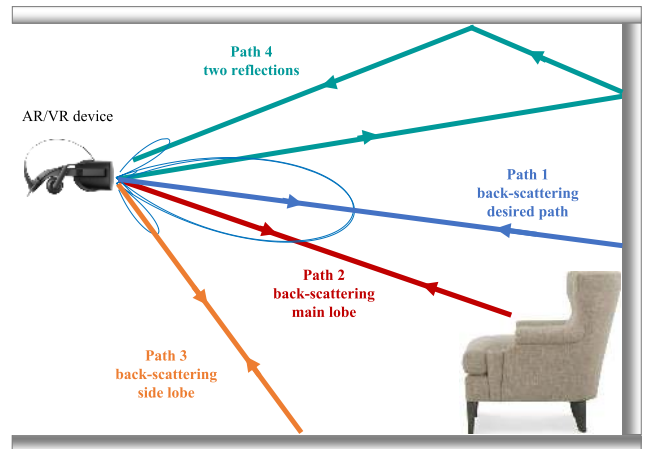
FIGURE 6. The multipath estimation challenge for scene range estimation is illustrated. The design challenge is how the sensing framework can detect and estimate the range through the desired channel path (path 1 in blue) and avoid making faulty estimation because of the other undesired paths (paths 2-4) in the environment.

of interest. For example, if the receiver estimates the range/depth based on a wrong path, this may noticeably degrade the accuracy of the depth map estimation. This challenge is depicted in Fig. 6. As illustrated, the challenge is how to design the sensing framework (the codebook and post-processing) to detect the desired channel path (the path in blue) while filtering out all the undesired channel paths. Examples of undesired paths are the paths 2-4. Path 2 is transmitted and received within the main lobe. Path 3 is transmitted and received within the side lobe. Path 4 experiences multiple reflections instead of back-scattering, before reaching back the receiver. It has to be noted that the diffuse scattering and specular reflection properties of the mmWave signals are still crucial for constructing depth maps despite their contribution to the inter-path interference challenge. Without these properties, the sensing framework may not be able to construct a meaningful depth scene of the surrounding environment.

In the next two sections, we efficiently design the two elements of our proposed depth map sensing framework, namely the sensing codebook and the post-processing, to address these challenges.

VI. DEPTH MAP BASED DESIGN FOR SENSING CODEBOOKS

As discussed in Section V-A, our objective is to design a sensing beamforming codebook that fits the rectangular grid of the depth camera. In this section, we first present our codebook design that achieves this objective. Then, we incorporate a new side-lobe reduction approach to ameliorate the inter-path interference problem.

A. PROPOSED CODEBOOK DESIGN

Since the objective from the beamforming-combining pair codebook design is for the codebook grid to match the desired

(a) The classical beamforming/combining codebook

(b) The proposed beamforming/combining codebook
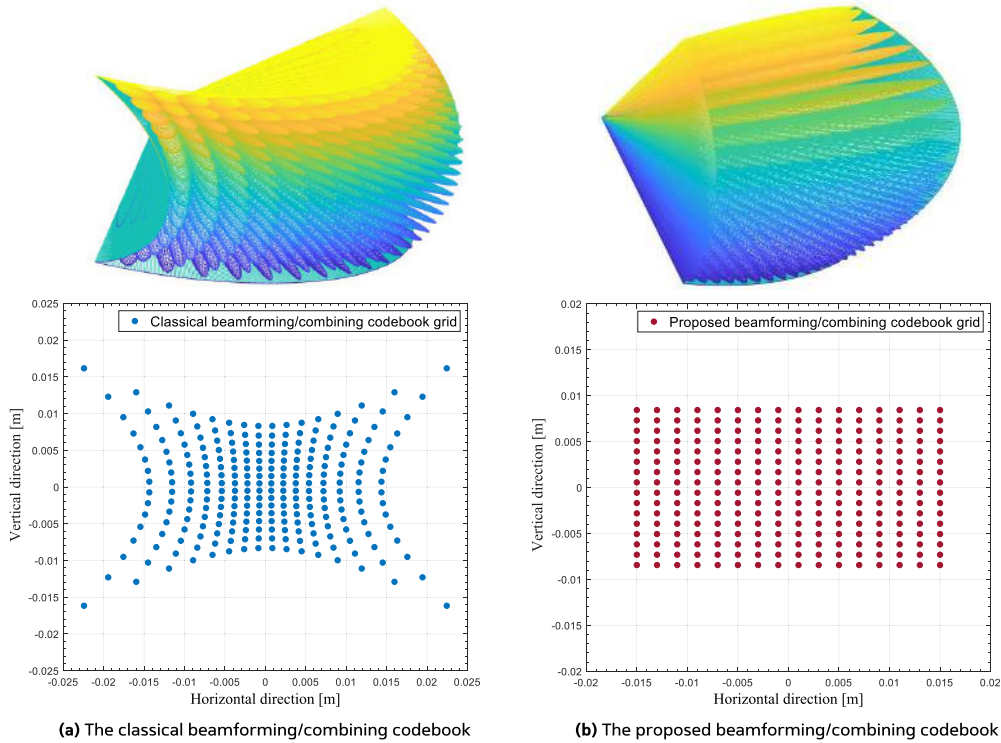
FIGURE 7. The comparison between (a) the classical (on the left side) and (b) the proposed (on the right side) beam codebook design is demonstrated for a scene of 100° field of view and 16/9 aspect ratio, using 16 × 16 UPAs. The proposed codebook eliminates any grid mismatch distortion. The top figures are the 3D codebook radiation patterns, while the bottom figures are the 2D codebook grids at a plane within 13.32mm depth.

rectangular grid of a range/depth scene, we start with the relevant camera geometry equations. The scene definition starts by defining the key quantities of the field of view, FoV, and the scene aspect ratio, A_R. Let the field of view be centered around the boresight antenna array direction. It is worth noting that the separation distance of the camera plane away from the antenna array reference point, aka the focal length, is irrelevant in our codebook design. This is based on the notice that the beamforming/combining codebook design normally depends on angles rather than distances.

In a general sense, for any chosen value of focal length, the sensor grid points' coordinates are first calculated to determine the codebook angles accordingly. More specifically, assume that the focal length is set to a certain value, F_L. The camera plane width, aka the sensor grid width in the horizontal dimension, S_H, and camera plane height, aka the sensor grid height in the vertical dimension, S_V, can be calculated as

$$S_H = 2F_L \tan(\text{FoV}/2), \text{ and } S_V = S_H/A_R. \quad (13)$$

For designing a beamforming-combining pair codebook, let $N = N_V \times N_H$, where N_V and N_H denote the number of UPA antennas on the elevation (vertical) and azimuth (horizontal) dimensions, respectively. Consider an oversampled beamforming codebook of $M = \overline{N}_V \overline{N}_H$ beams, where $\overline{N}_V = N_V F_V^{OS}$ and $\overline{N}_H = N_H F_H^{OS}$ with F_V^{OS} and F_H^{OS} denoting the oversampling factors in the elevation and azimuth

dimensions, respectively. The grid spacing in the vertical and horizontal directions are expressed as $Q_V = S_V/\overline{N}_V$ and $Q_H = S_H/\overline{N}_H$. Notice that the codebook resolution of $\overline{N}_V \overline{N}_H$ beams will be mapped at the end to the desired up-scaled depth image resolution of $M_{res} = M_h \times M_w$ pixels.

Let the x- and z-axes be aligned in the direction of the sensor grid width and height respectively, and let the y-axis be the direction of the depth. The (x, y, z) rectangular coordinates of the sensor grid points on the camera plane can then be defined as

$$(x, y, z) \in \mathcal{C}, \quad \mathcal{C} = \mathcal{X} \times \mathcal{Y} \times \mathcal{Z},$$
$$\mathcal{X} = \left\{ x : x \in \left\{ \frac{-S_H}{2} + \frac{Q_H}{2}, \frac{-S_H}{2} + \frac{3Q_H}{2}, \dots, \frac{S_H}{2} - \frac{Q_H}{2} \right\} \right\},$$
$$\mathcal{Y} = \{ y : y = F_L \},$$
$$\mathcal{Z} = \left\{ z : z \in \left\{ \frac{-S_V}{2} + \frac{Q_V}{2}, \frac{-S_V}{2} + \frac{3Q_V}{2}, \dots, \frac{S_V}{2} - \frac{Q_V}{2} \right\} \right\},$$

(14)

where we note that $|\mathcal{C}| = \overline{N}_V \overline{N}_H = M$. After defining the (x, y, z) coordinates of every grid point on the camera plane, their M corresponding (θ_z, θ_x) angles with respect to the z- and x-axes can now be calculated using the mapping from rectangular to spherical coordinates, such that

$$\mathcal{O} = \left\{ (\theta_z, \theta_x) : \theta_z = \left[\frac{\pi}{2} - \arctan\left(\frac{z}{\sqrt{x^2 + y^2}} \right) \right], \right.$$
$$\left. \theta_x = \left[\frac{\pi}{2} - \arctan\left(\frac{x}{\sqrt{y^2 + z^2}} \right) \right], (x, y, z) \in \mathcal{C} \right\}. \quad (15)$$

Finally, after calculating the (θ_z, θ_x) angles for each and every grid point, the beamforming codebook, $\mathcal{F}$, for an $N_H \times N_V$ transmit UPA, is then expressed as

$$\mathcal{F} = \left\{ \mathbf{f} \in \mathbb{C}^{N \times 1} : \mathbf{f} = \widetilde{\mathbf{b}}_V(\theta_z) \circ \widetilde{\mathbf{b}}_H(\theta_x), (\theta_z, \theta_x) \in \mathcal{O} \right\},$$

$$\widetilde{\mathbf{b}}_V(\theta_z) = \left[1, e^{-j\kappa d_s \cos(\theta_z)}, \ldots, e^{-j(N_V-1)\kappa d_s \cos(\theta_z)} \right]^T,$$

$$\widetilde{\mathbf{b}}_H(\theta_x) = \left[1, e^{-j\kappa d_s \cos(\theta_x)}, \ldots, e^{-j(N_H-1)\kappa d_s \cos(\theta_x)} \right]^T,$$

(16)

where $\kappa = \frac{2\pi}{\lambda}$ is the wave number, λ is the operating wavelength, and d_s is the antenna element spacing between adjacent UPA elements in meters. $\widetilde{\mathbf{b}}_H \in \mathbb{C}^{N_H \times 1}$ and $\widetilde{\mathbf{b}}_V \in \mathbb{C}^{N_V \times 1}$ are the horizontal and vertical basic vectors used for constructing the beamforming codebook. We will call these vectors, $\widetilde{\mathbf{b}}_H$ and $\widetilde{\mathbf{b}}_V$, the *constituent* horizontal and vertical beamforming vectors, respectively. In our depth estimation problem, the receive combining codebook, $\mathcal{W}$, can be similarly defined for the $N_H \times N_V$ receive UPA. For such case, the cardinalities of the sets are equal, $|\mathcal{W}| = |\mathcal{F}| = |\mathcal{C}| = |\mathcal{O}| = M$. Further, let $\mathbf{F} \in \mathbb{C}^{N \times M}$ and $\mathbf{W} \in \mathbb{C}^{N \times M}$ be the matrices that consist of the codebooks beams of $\mathcal{F}$ and $\mathcal{W}$. Then, the proposed sensing beamforming-combining pair codebook $\mathcal{P}$ can be expressed as

$$\mathcal{P} = \left\{ (\mathbf{f}_m, \mathbf{w}_m) \in \mathbb{C}^{N \times 1} \times \mathbb{C}^{N \times 1} : \mathbf{f}_m = [\mathbf{F}]_{:,m}, \right.$$

$$\left. \mathbf{w}_m = [\mathbf{W}]_{:,m}, m \in \{1, \ldots, M\} \right\}. \quad (17)$$

A comparison between the classical and the proposed beam codebook design is demonstrated in Fig. 7 for a scene of 100° field of view and 16/9 aspect ratio, using 16×16 UPAs. The top figures are the 3D codebook radiation patterns, while the bottom figures are the 2D codebook grids at a plane within 13.32mm depth. As shown, the proposed beam codebook eliminates any grid mismatch distortion.

B. SIDELOBE REDUCTION APPROACH

As discussed in Section V-B, to rectify the inter-path interference problem, the sensing framework needs to filter out the undesired channel paths. As illustrated in Fig. 6, one type of undesired channel paths is the type of paths transmitted from/received by the sidelobes of a codebook beam. For this reason, we propose an efficient sidelobe reduction (SLR) approach. In [41], [42], an SLR approach was proposed for low sidelobe beamforming in uniform circular arrays. Inspired by their work, we propose a new efficient sidelobe reduction approach to uniform planar arrays (UPAs) to reduce beamforming/combining sidelobe levels.

The key idea of this approach is when applying different weights on the beamforming/combining vector elements, the sensing framework can control the beam radiation pattern in a way to increase the power difference between the mainlobe and the sidelobes. Specifically, let $\mathbf{c}_H \in \mathbb{R}^{N_H \times 1}$ and $\mathbf{c}_V \in \mathbb{R}^{N_V \times 1}$ represent the horizontal and vertical weight vectors for sidelobe reduction. Let $\mathbf{b}_H$ and $\mathbf{b}_V$ denote the

horizontal and vertical constituent beamforming vectors after sidelobe reduction. The updated beamforming codebook, $\mathcal{F}$, for an $N_H \times N_V$ transmit UPA, can then be rewritten as

$$\mathcal{F} = \left\{ \mathbf{f} \in \mathbb{C}^{N \times 1} : \mathbf{f} = \mathbf{b}_V(\theta_z) \circ \mathbf{b}_H(\theta_x), (\theta_z, \theta_x) \in \mathcal{O} \right\},$$

$$\mathbf{b}_V(\theta_z) = \widetilde{\mathbf{b}}_V(\theta_z) \odot \mathbf{c}_V, \ \mathbf{b}_H(\theta_x) = \widetilde{\mathbf{b}}_H(\theta_x) \odot \mathbf{c}_H,$$

$$[\mathbf{c}_V]_{r_V} = e^{\frac{-(r_V - \mu_V)^2}{2\sigma_V^2}}, \mu_V = \frac{N_V}{2}, \sigma_V = \frac{N_V}{\delta_V}, r_V \in \{1, \ldots, N_V\},$$

$$[\mathbf{c}_H]_{r_H} = e^{\frac{-(r_H - \mu_H)^2}{2\sigma_H^2}}, \mu_H = \frac{N_H}{2}, \sigma_H = \frac{N_H}{\delta_H}, r_H \in \{1, \ldots, N_H\},$$

(18)

where δ_H, δ_V denote the sidelobe reduction control variables; the higher the values, the greater reduction in the sidelobe power levels compared to the mainlobe power level. The updated combining codebook $\mathcal{W}$ can be similarly defined. The beam codebook $\mathcal{P}$ follows the same definition in (17).

Fig. 8 illustrates the radiation pattern in dB for one beamforming vector out of the updated beamforming codebook, $\mathcal{F}$, for different values of the sidelobe reduction control variables, δ_H, δ_V. As depicted, increasing the values of the control variables increases the power gap between the mainlobe level and the sidelobes levels. To take into consideration the phase quantization of the RF phase shifters in the AR/VR transceiver architecture previously shown in Fig. 2, we examine the effect of 2-bit phase quantization on the power radiation pattern. The 2-bit discrete phase shift set is $\left\{ 0, \frac{\pi}{2}, \pi, \frac{3\pi}{2} \right\}$. Fig. 9 compares the normalized power radiation pattern between the case of continuous phase shifts and the case of 2-bit quantized phase shifts. As depicted, the phase quantization affects the beam pattern shape of the sidelobes.
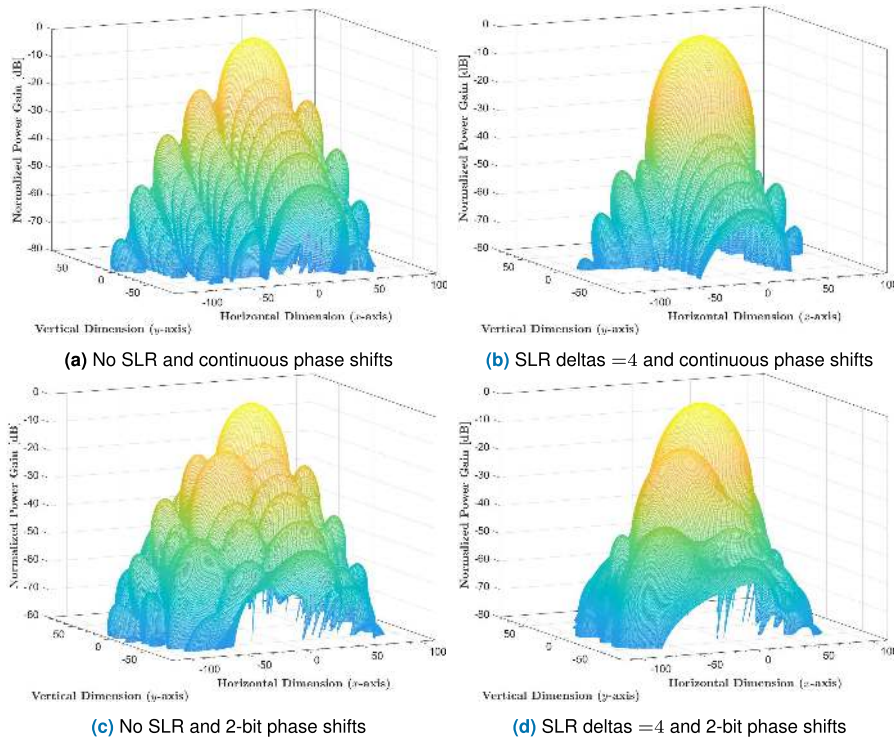
One main advantage of this approach is its computational efficiency; as formulated, only two element-wise multiplication between the weight vectors and the constituent beamforming vectors, $\widetilde{\mathbf{b}}_H, \widetilde{\mathbf{b}}_V$, are needed to update the beam radiation pattern. This multiplication, however, requires an analog beamforming architecture with the capability of changing both the phase and magnitude. In the results section, we only used this SLR-based beam codebook in the simulations of Fig. 24. In the future work, it is interesting to explore phase-only approximations of this SLR-based beam codebook structure. By contrast, reducing the sidelobe levels dramatically increases the beamwidth of the mainlobe, as depicted in Fig. 8. The increased mainlobe beamwidth, however, can be mitigated by the other solutions proposed for rectifying the inter-path interference problem, e.g. the successive interference cancellation (SIC) algorithm and the joint processing (JP) solution, as will be described in the following section.

VII. PROPOSED SCENE RANGE/DEPTH ESTIMATION

In this section, given a pre-designed beamforming/combining codebook, $\mathcal{P}$, we propose an efficient approach for the scene range/depth estimation in AR/VR devices. As depicted in Fig. 4, once the beamforming/combining codebook has been designed, the AR/VR transmits the sensing signal

(a) No SLR   **(b)** SLR, deltas = 3   **(c)** SLR, deltas = 4
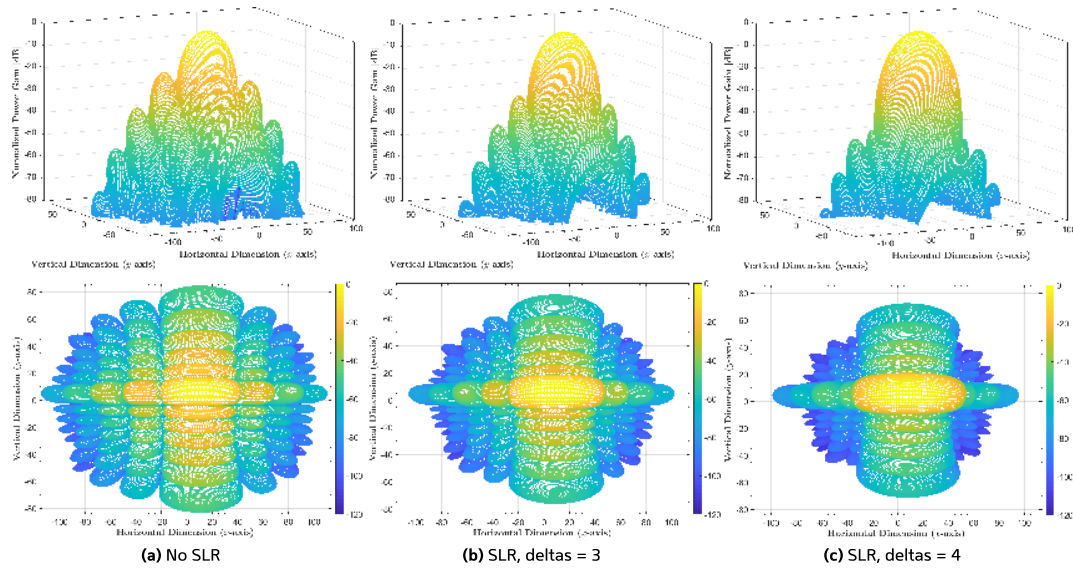
FIGURE 8. Normalized power radiation pattern comparison between (a) the case without the sidelobe reduction (SLR) approach, (b) the case with the SLR approach where $\delta_H = \delta_V = 3$, and (c) where $\delta_H = \delta_V = 4$. As shown, increasing the values of the control variables (the deltas) increases the gap between the mainlobe level and the sidelobes levels. The top figures are the 3D views of the patterns while the bottom figures are the top views.

(a) No SLR and continuous phase shifts   **(b)** SLR deltas =4 and continuous phase shifts

(c) No SLR and 2-bit phase shifts   **(d)** SLR deltas =4 and 2-bit phase shifts

FIGURE 9. Normalized power radiation pattern comparison between the case with no phase quantization and the case with 2-bit phase quantization, for two scenarios: without or with the sidelobe reduction (SLR) approach where $\delta_H = \delta_V = 4$.

while sweeping over all the beamforming-combining vector pairs. Specifically, for a beamforming-combining vector pair $(\mathbf{f}_m, \mathbf{w}_m)$, where $m \in \{1, \ldots, M\}$, the receive *sensing* signal, $\mathbf{y}_m \in \mathbb{C}^{NP+L_{\mathsf{d}}}$, can be modeled as in (6) and (7). After reception, the acquired sensing signals are processed to estimate the range and depth maps, as will be thoroughly explained in this section. Our proposed post-processing solution has three main elements: (i) The use of oversampled/overlapped beams, (ii) the successive interference cancellation based management of inter-target and inter-path interference, and (iii) the joint processing of the signals received using the codebook beams to realize high-resolution

---

Algorithm 1 Successive Interference Cancellation

Inputs: Receive sensing signal $y_m[n]$, transmit preamble signal $s^{\mathrm{p}}[n]$, threshold level A_{TH}, beamforming-combining pair codebook $\mathcal{P}$.

Outputs: Candidate delay set for each beam, $\mathcal{T}_m$, $\forall m \in \{1, \ldots, M\}$,.

1: **for** $m = 1$ **to** M **do**

2:     **Initialize:** Updated signal $\tilde{y}_m[n] \leftarrow y_m[n]$, solution set $\mathcal{T}_m \leftarrow \emptyset$, $\forall m$, and $\tilde{A} \leftarrow A_{\mathrm{TH}}$.

3:     **while** $\tilde{A} \geq A_{\mathrm{TH}}$ **do**

4:         Calculate the delay of the path with maximum cross-correlation

$$\tilde{q} \leftarrow \arg\max_{q:q \in \mathcal{Q}} \left| \sum_{n=L_{\mathrm{y}}-N_{\mathrm{y}}}^{L_{\mathrm{y}}-1} s^{\mathrm{p}}[n] \times (\tilde{y}_m[n-q])^* \right|^2 .$$

5:         Add the candidate delay $\tilde{q}$ to the solution set

$$\mathcal{T}_m \leftarrow \mathcal{T}_m \cup \{\tilde{q}\} .$$

6:         Calculate the energy of the transmit signal up to $\tilde{q}$

$$E_{\mathrm{Q}} \leftarrow \sum_{n=L_{\mathrm{y}}-N_{\mathrm{y}}}^{L_{\mathrm{y}}-\tilde{q}-1} \left| s^{\mathrm{p}}[n] \right|^2 .$$

7:         Perform interference cancellation at $\tilde{q}$

$$\tilde{A} \leftarrow \left[\sum_{n=L_{\mathrm{y}}-N_{\mathrm{y}}}^{L_{\mathrm{y}}-1} s^{\mathrm{p}}[n] \times (\tilde{y}_m[n-\tilde{q}])^* \right] .$$

$$\tilde{y}_m[n] \leftarrow \tilde{y}_m[n] - \frac{\tilde{A}}{E_{\mathrm{Q}}} s^{\mathrm{p}}[n-\tilde{q}].$$

8:         Calculate the next max. in the cross-correlation

$$\tilde{A} \leftarrow \max_{q:q \in \mathcal{Q}} \left| \sum_{n=L_{\mathrm{y}}-N_{\mathrm{y}}}^{L_{\mathrm{y}}-1} s^{\mathrm{p}}[n] \times (\tilde{y}_m[n-q])^* \right|^2 .$$

---

and accurate depth maps. Next, we explain these three elements in Sections VII-A-VII-C before presenting the scene range/depth map construction approach in Section VII-D.

A. OVERLAPPED BEAMS

With the objective of increasing the resolution of the mmWave MIMO based depth maps, we propose to adopt oversampled sensing codebooks to scan the surrounding environment. In particular, for the sensing codebook, we adopt the developed codebook in Section VI-A with oversampling factors of $F_{\mathrm{H}}^{\mathrm{OS}}$ and $F_{\mathrm{V}}^{\mathrm{OS}}$ in the azimuth and elevation directions. While the oversampled codebook has the potential of enhancing the depth map resolution, it is important to note that advanced post-processing (for the receive signals using these oversampled beams) needs to be incorporated to achieve this goal. The reason mainly goes back to the wide beamwidth

(and low spatial resolution) of the codebook beams, which is fundamentally limited by the number of AR/VR antennas. This wide beamwidth leads to a number of challenges: (i) The spatial regions scanned by the oversampled beams have high overlap. This makes it hard to differentiate between the depths of the different objects in the depth map pixels, which challenges the objective of realizing high-resolution depth maps. (ii) Since the codebook beams still have wide beamwidth, the inter-target interference problem discussed in Section V-B still exists.

To address these challenges, we propose a novel post-processing approach based on successive interference cancellation and joint-beam processing. This approach in summarized in two main steps as follows. In the first step, a successive interference cancellation (SIC) based algorithm is used to detect the most dominant channel paths contributing to the range/depth estimation of the region covered by each codebook beam. These paths form a set of candidate ranges/depths for the scene range/depth estimation. In the second step, a developed joint-beam processing solution selects one range/depth out of the set of candidate ranges/depths formed by the SIC algorithm. These two sequential algorithms are discussed in detail in the following two subsections.
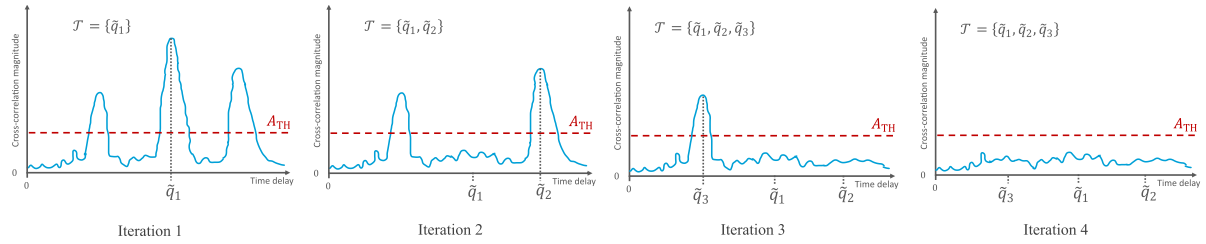
B. SUCCESSIVE INTERFERENCE CANCELLATION

The main goal of the successive interference cancellation (SIC) algorithm is to detect all the dominant paths that might contribute to the range estimation of the region of interest. This is motivated by its good performance in multi-target detections problems [43]. The SIC algorithm is applied in the discrete-time domain and is summarized in Algorithm 1. The algorithm is described as follows. Let the length of the receive sensing sequence $y_m[n]$ be $L_{\mathrm{y}} = N^{\mathrm{p}} + L_{\mathrm{d}}$ symbols. First, as shown in Fig. 10, for every codebook beam, the delay position of the maximum cross-correlation magnitude value is detected. $\mathcal{Q}$ is the set of possible delays. Second, the SIC algorithm encodes the transmit preamble signal to be shifted to this delay position and subtracted it from the received signal. Afterwards, the algorithm repeats itself to detect the second local maximum above the threshold value. Finally, The SIC algorithm stops iterating when all the local maxima above the threshold value are detected. The output of this algorithm is a set of candidate delays for every codebook beam. These sets pass as input to the next algorithm, the joint processing solution, as will be explained in the next part. In Fig. 10, note that the cross-correlation magnitude plot appears to be drawn as a continuous plot, only for illustration purposes. The actual cross-correlation magnitude, however, is expressed in discrete time delays.
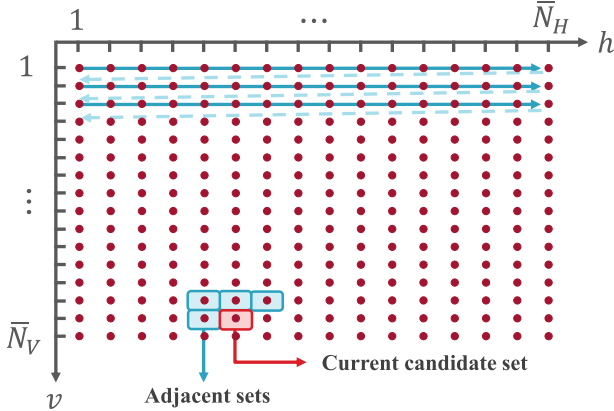
C. JOINT PROCESSING SOLUTION

The purpose of the joint processing (JP) solution between the overlapped beams is to estimate the transitions in depth/range maps more accurately. The proposed JP solution is summarized in Algorithm 2. The algorithm is described in detail as

FIGURE 10. The operation of the successive interference cancellation (SIC) algorithm is illustrated. The delay position of the maximum cross-correlation is first detected. The SIC algorithm then encodes a signal shifted at this delay position and subtracted it from the receive signal. After that, the algorithm repeats itself until all the local maxima above the threshold value are detected.

FIGURE 11. This figure illustrates the basic operation of the joint processing (JP) solution for overlapped beams. The JP solution sweeps from left to right, then from top to bottom. The JP solution decides on which path to choose from the current candidate set by a simple comparison with the sets of the surrounding grid points.

follows. First, the JP solution works on the candidate delay sets, the output from the SIC algorithm, $\{\mathcal{T}_m\}_{m=1}^M$, to choose one range estimate out of the candidate delay set. This processing, however, is employed relative to the 2D codebook grid, as illustrated in Fig. 11. Following this notion, the linear indices in $\mathcal{T}_m$ is now converted into matrix subscripts $\mathcal{T}_{h,v}$ through the transformation $m = (v - 1)\overline{N}_H + h$, such that $\mathcal{T}_m = \mathcal{T}_{h,v}$, where v is the elevation beam index (vertical grid index) and h is the azimuth beam index (horizontal grid index). The objective is to calculate the scene range estimates across all beam directions, $\left(\hat{\rho}_{h,v}\right)^{\text{SRE}}$, $\forall h, v$.

As shown in Fig. 11, the JP solution sweeps from left to right, then from top to bottom. For each grid point, the JP solution uses (i) the set of the current grid point, named as the "current set", and (ii) the sets of the previous adjacent grid points to construct a "common adjacent set". This common adjacent set is the union of the sets of all previous adjacent grid points. Then, to investigate if a new object/surface transition appears, this current set is compared with the common adjacent set to detect if there is any set difference. This is based on the notion that the difference set can probably be the new edges that will appear in the range map while sweeping. If the set difference is not empty, then the solution chooses the path with the least time-of-flight from the set difference. Otherwise, if the set difference is empty, then the solution chooses the path with the least time-of-flight from the current set.

Algorithm 2 Joint Processing Solution

Inputs: Candidate delay set for each beam, $\mathcal{T}_{h,v}, \forall h \in \{1, \ldots, \overline{N}_H\}, \forall v \in \{1, \ldots, \overline{N}_V\}$.

Outputs: Scene range estimate $\left(\hat{\rho}_{h,v}\right)^{\text{SRE}}$, $\forall h, v$.

1: **Initialize:** Common adjacent set $\mathcal{N}_{h,v} \leftarrow \emptyset, \forall h, v$, difference set $\mathcal{M}_{h,v} \leftarrow \emptyset, \forall h, v$.

2: **for** $v = 1$ **to** $\overline{N}_V$ **do**

3:    **for** $h = 1$ **to** $\overline{N}_H$ **do**

4:       Construct the common adjacent set

$$\mathcal{N}_{h,v} \leftarrow \left(\mathcal{T}_{h-1,v} \cup \mathcal{T}_{h,v-1} \cup \mathcal{T}_{h-1,v-1} \cup \mathcal{T}_{h+1,v-1}\right).$$

5:       Construct the difference set

$$\mathcal{M}_{h,v} \leftarrow \mathcal{T}_{h,v} \setminus \mathcal{N}_{h,v}.$$

6:       **if** $\mathcal{M}_{h,v} \neq \emptyset$ **then**

7:          Choose the least delay from the difference set

$$\left(\hat{\rho}_{h,v}\right)^{\text{SRE}} \leftarrow \frac{\varsigma T_S}{2} \min \mathcal{M}_{h,v}.$$

8:       **else**

9:          Choose the least delay from the candidate set

$$\left(\hat{\rho}_{h,v}\right)^{\text{SRE}} \leftarrow \frac{\varsigma T_S}{2} \min \mathcal{T}_{h,v}.$$

D. RANGE/DEPTH MAP CONSTRUCTION

In this section, we formulate the depth map construction approach, the last step in Fig. 4. In summation of the broader view, the mmWave MIMO sensing based range/depth map estimation framework is outlined in Algorithm 3. The algorithm steps are summarized as follows. *Step 1* refers to the design of the beamforming-combining pair codebook $\mathcal{P}$ was covered in Section VI. *Step 4* refers to the successive interference cancellation described in Section VII-B. *Step 5* refers to the joint processing solution detailed in Section VII-C. After that, in *Step 6*, the fine range estimate can be calculated, such that $\left(\hat{\rho}_m\right)^{\text{MC}} = \left(\hat{\rho}_m\right)^{\text{SRE}} + \left(\hat{\rho}_m\right)'$, where $\left(\hat{\rho}_m\right)'$ is computed from the algorithm described in Section IV-B2. Next, after calculating the range estimates, the upcoming steps (*Steps 7,8*) are focused on constructing the range and depth maps. Note that the range of an object is actually the radial distance in spherical coordinates. Fortunately, the (x, y, z) rectangular coordinates of the sensor grid points on the

Algorithm 3 mmWave MIMO Sensing Based Range/Depth Estimation Framework

---

Inputs: Field of view FoV, aspect ratio A_R, number of horizontal and vertical beams $\overline{N}_H$, $\overline{N}_V$.

Outputs: Range map $\widehat{\mathbf{R}}_{map}$, depth map $\widehat{\mathbf{D}}_{map}$.

1: Design the beamforming-combining pair codebook, $\mathcal{P}$, following Section VI.
2: **for** $m = 1$ **to** M **do**   ▷ For each pair $(\mathbf{f}_m, \mathbf{w}_m)$.
3:     Acquire receive *sensing* signal, $y_m[n]$, as in (6), $\forall n \in \{0, 1, \dots, N^p + L_d - 1\}$.
4:     Calculate the candidate delay set for each beam, $\mathcal{T}_m$, $\forall m$, as in Algorithm 1.
5:     Calculate the scene range estimate, $\left(\hat{\rho}_m\right)^{SRE}$, $\forall m$, as in Algorithm 2.
6:     Calculate fine range estimates, following Section IV-B2.

$$\left(\hat{\rho}_m\right)^{MC} \leftarrow \left(\hat{\rho}_m\right)^{SRE} + \left(\hat{\rho}_m\right)', \forall m.$$

7: Construct the range map, $\widehat{\mathbf{R}}_{map}$, from (20).
8: Construct the depth map, $\widehat{\mathbf{D}}_{map}$, from (21).

---

camera plan were already calculated for the design of the beamforming-combining pair codebook using (14). These rectangular coordinates in (14) can then be converted to spherical coordinates, such that

$$\mathcal{S} = \left\{ (\theta_z, \Phi) : \theta_z = \left[\frac{\pi}{2} - \arctan\left(\frac{z}{\sqrt{x^2 + y^2}} \right) \right], \right.$$
$$\left. \Phi = \arctan\left(\frac{y}{x} \right), (x, y, z) \in \mathcal{C} \right\}. \quad (19)$$

In order to construct the matrices for the range and depth maps, let $\mathbf{\Theta}, \mathbf{\Phi} \in \mathbb{R}^{\overline{N}_V \times \overline{N}_H}$ be the matrices that represent the angles of the spherical coordinates $(\theta_z, \Phi) \in \mathcal{S}$, respectively. Following *Step 6* in Algorithm 3, the range map estimate $\widehat{\mathbf{R}}_{map} \in \mathbb{R}^{\overline{N}_V \times \overline{N}_H}$ can be expressed as

$$\left[\widehat{\mathbf{R}}_{map}\right]_{v,h} = \left(\hat{\rho}_m\right)^{MC}, \quad m = (v - 1)\overline{N}_H + h, \quad (20)$$

where $m \in \{1, \dots, M\}$, $h \in \{1, \dots, \overline{N}_H\}$, $v \in \{1, \dots, \overline{N}_V\}$. Given the angles in spherical coordinates and the range map estimate, the depth map estimate $\widehat{\mathbf{D}}_{map} \in \mathbb{R}^{\overline{N}_V \times \overline{N}_H}$ can then be expressed as

$$\left[\widehat{\mathbf{D}}_{map}\right]_{v,h} = \left| \hat{\rho} \sin(\theta_z) \sin(\Phi) \right|, \quad (21)$$

where $\hat{\rho} = \left[\widehat{\mathbf{R}}_{map}\right]_{v,h}$, $\theta_z = [\mathbf{\Theta}]_{v,h}$, $\Phi = [\mathbf{\Phi}]_{v,h}$, $\forall v, h$.

Finally, since the range and depth map resolutions are set to $\overline{N}_H \times \overline{N}_V$, two-dimensional image interpolation can be employed to scale the maps to the desired resolutions, $M_h \times M_w$. Examples of interpolation methods are the nearest neighbor interpolation and the bicubic interpolation. Although the bicubic interpolation can probably be the interpolation method of choice for achieving more estimation accuracy, the nearest neighbor interpolation is more computationally efficient. In the simulation results of Section VIII, we evaluate the two interpolation approaches for our mmWave MIMO based depth map construction problem.

VIII. SIMULATION RESULTS

In this section, we evaluate the performance of the proposed mmWave based depth estimation approach. First, we describe the adopted simulation framework in Section VIII-A before extensively studying the estimation accuracy of the proposed approach under various scenarios and system parameters. The simulation results presented can be of great usefulness for various applications; they can be generally applied to AR/VR devices, smart home devices, or auto drive devices.

A. SIMULATION FRAMEWORK

Since the depth estimation heavily depends on the environment under test, it is crucial to evaluate the performance of the proposed solution based on realistic channels. This motivates using channels generated by accurate ray-tracing to capture the sensing dependence on the environment geometry, scatterers' materials, AR/VR position, etc. This is why we designed the simulations models using Remcom Wireless InSite [16], which is an accurate 3D ray-tracing simulator. Further, to efficiently incorporate diffuse scattering models, we need to have highly detailed floor plans with a sufficient number of faces. To achieve this objective, we resorted to the high-fidelity game engine, Blender [17], to build accurate floor plans. These plans/models are then exported to Wireless InSite to obtain the ray-tracing outputs, and finally to MATLAB to construct the channel models in (4) and implement the proposed depth estimation approach. The proposed evaluation framework is illustrated in Fig. 12. For benchmarking, we also use the Blender floor plans to obtain the ground truth depth maps, which are essential to evaluate the accuracy of our solutions. The ground truth maps are generated by placing a Blender camera at the same position of the UPA reference antenna element, and adjusting the Blender camera parameters to capture the same field of view.

1) SIGNAL MODEL

We adopt the signal model described in Section II with a focus on the sensing system performance. The AR/VR device is assumed to be fixed in position. Unless otherwise mentioned, the UPA size is 16×16 antennas ($N_H = N_V = 16$) at the mmWave 60GHz operating band with transmission bandwidth of 2GHz. The antenna elements have a gain of 0dBi with half-wavelength antenna spacing. The transmit power is set to 30dBm. The preamble sequence is the same as the one in the single carrier PHY packet preamble of the IEEE 802.11ad standard (3328 symbols). M preamble sequences are used to sense the environment via M beamforming-combining pairs. For the sake of calculating a rough estimate of the time allocated for environment sensing through transmission and reception, assume that all the M preamble sequences are transmitted sequentially with guard intervals in between. The highest M value reported in the upcoming simulation results is 4096 beams. Assuming a sampling rate of 2Gsps, the estimate of the longest sensing time is then ≈ 7ms.
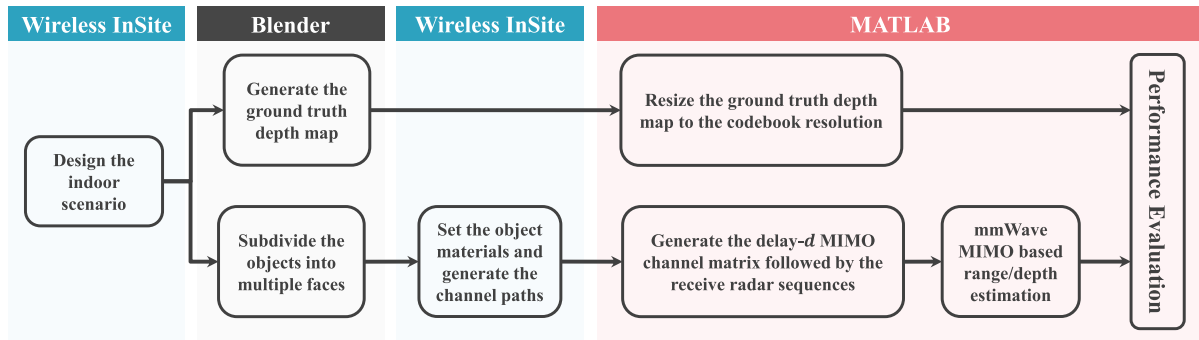
FIGURE 12. This figure demonstrates the adopted simulation framework for scene depth estimation. The framework consists of designing the indoor setup, generating the ground truth range/depth maps, and constructing the estimated maps for performance evaluation. For more complex setups, designing the indoor scenarios jointly in Wireless InSite and Blender can be more effective.

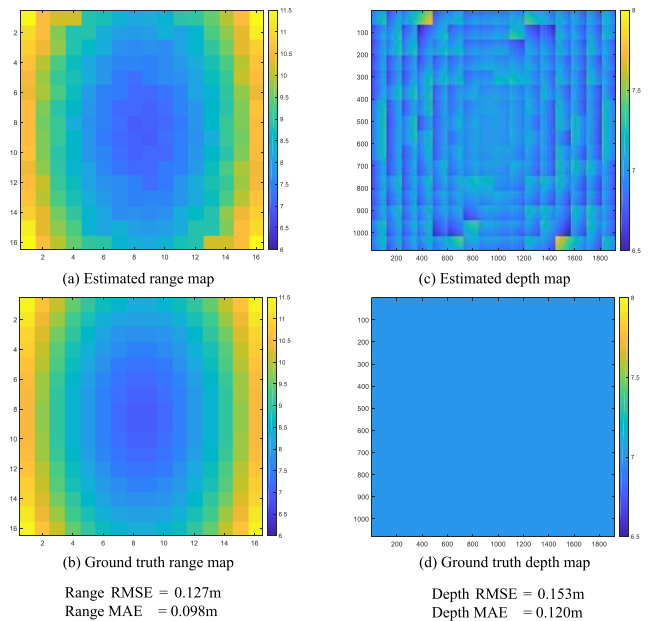**TABLE 1.** The adopted diffuse scattering parameters for different materials.

Diffuse Scattering Parameter	Concrete	Ceilingboard	Wood	Floorboard	Drywall	Glass
Scattered to incident electric field ratio	40%	30%	15%	15%	10%	0%
Forward to backward scattering power ratio	75%	75%	75%	75%	75%	75%
Cross-polarization ratio	40%	40%	40%	40%	40%	40%
Narrowness of the scattering lobes	40%	40%	40%	40%	40%	40%

2) CHANNEL GENERATION

The channel matrix, $\mathbf{H}_d$, is generated in two steps. The first step is generating the channel rays using the ray-tracing software, Wireless InSite. The Wireless InSite propagation model is set to 'X3D' with 0.1° ray-spacing and enabled mode of diffuse scattering. Up to three reflections, one diffraction, and one transmission properties are allowed for each ray in the Wireless InSite simulation. The diffuse scattering model used is "directive with backscatter"; this model is fixed across all materials in all the testing scenarios. The chosen diffuse scattering model creates two scattering lobes; a forward lobe of diffuse scattered power centered on the direction of specular reflection and a backward lobe centered on the opposite direction of incidence. The diffuse scattering parameters of the different materials are summarized in Table 1. The values reported in Table 1 follow the ITU default parameter values at 60GHz. The second step in the sensing channel generation is calculating the delay-d channel matrix out of the channel paths using the DeepMIMO dataset generation code [44]. Using these channels and following (4)-(3), the noisy receive sensing sequences are generated. The noise power is calculated based on a 2GHz bandwidth and a receiver noise figure of 7dB.
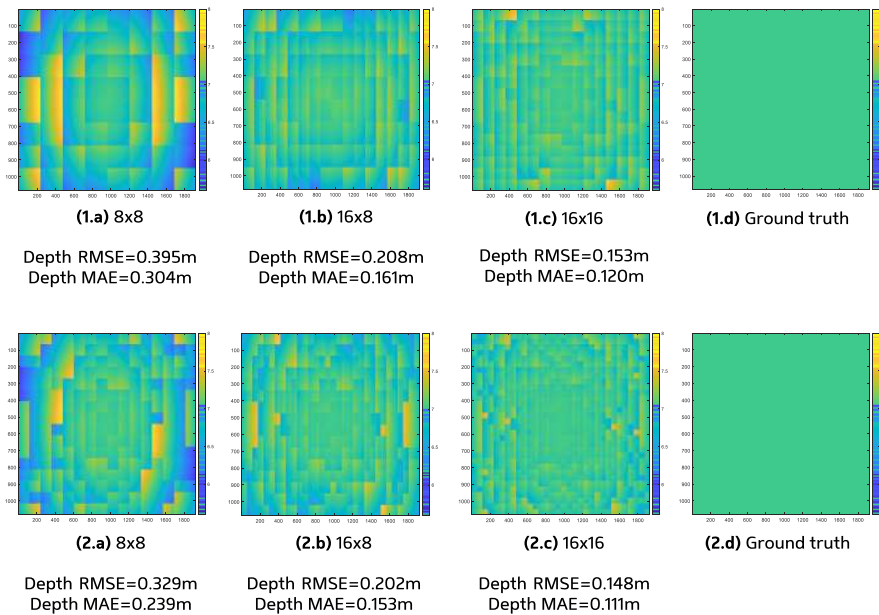
3) mmWave bASED DEPTH ESTIMATION PARAMETERS

The beamforming-combining pair codebook is designed based on a 100° field of view centered on the antenna array boresight, a 16/9 scene aspect ratio, and horizontal and vertical oversampling factors of unity. The ground truth depth maps are generated from Blender using a Blender camera with a 100° field of view, a focal length of 13.43mm

(a) Estimated range map

(c) Estimated depth map

(b) Ground truth range map

(d) Ground truth depth map

Range RMSE = 0.127m
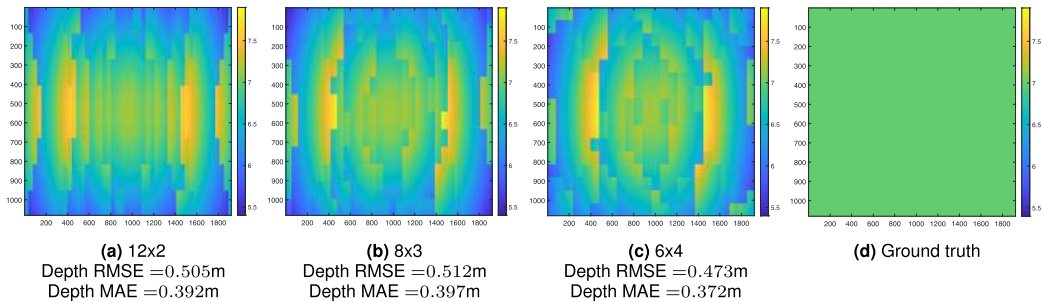Range MAE = 0.098m

Depth RMSE = 0.153m
Depth MAE = 0.120m

FIGURE 13. The maps for the one wall scenario are depicted for a separation distance of 7 meters from the AR/VR device with 16 × 16 UPAs. The depicted maps are the estimated maps (at the top), ground truth maps (at the bottom), range maps (on the left side), and 1080p depth maps (on the right side). Comparing (a) with (b), the range map estimation error: MAE = 0.098m. Comparing (c) with (d), the depth map estimation error: MAE = 0.12m.
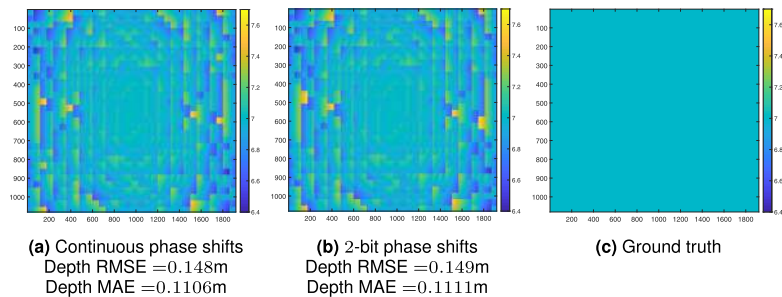
corresponding to a sensor width of 32mm. The ground truth depth map image quality is set to 1080p resolution; i.e., 1920 × 1080 pixels. Concerning the massive correlator, f_{est} is set to 100 multiple of the sampling frequency f_S; i.e., $\delta = \frac{f_{est}}{2f_S} = 50$. Unless mentioned otherwise, the

| (1.a) 8x8 | (1.b) 16x8 | (1.c) 16x16 | (1.d) Ground truth |

Depth RMSE=0.395m   Depth RMSE=0.208m   Depth RMSE=0.153m
Depth MAE=0.304m   Depth MAE=0.161m   Depth MAE=0.120m

| (2.a) 8x8 | (2.b) 16x8 | (2.c) 16x16 | (2.d) Ground truth |

Depth RMSE=0.329m   Depth RMSE=0.202m   Depth RMSE=0.148m
Depth MAE=0.239m   Depth MAE=0.153m   Depth MAE=0.111m

FIGURE 14. The depth maps for the one wall scenario are depicted for different antenna configurations and codebook resolutions, for a separation distance of 7 meters. Figures (a), (b), and (c) illustrate the estimated 1080p maps for 8 × 8, 16 × 8, and 16 × 16 UPAs. Figures (d) illustrate the ground truth maps. The top maps are with no codebook oversampling while the bottom maps are with codebook oversampling factors of two.

| (a) 12x2 | (b) 8x3 | (c) 6x4 | (d) Ground truth |

Depth RMSE =0.505m   Depth RMSE =0.512m   Depth RMSE =0.473m
Depth MAE =0.392m   Depth MAE =0.397m   Depth MAE =0.372m

FIGURE 15. The 1080p depth maps for the one wall scenario are depicted at different antenna configurations, for a separation distance of 7 meters. The same number of antenna elements is used (24 elements) and codebook oversampling factors of four are employed. Figures (a), (b), and (c) illustrate the estimated maps for 12 × 2, 8 × 3, and 6 × 4 UPAs. Figures (d) illustrate the ground truth depth map.

| (a) Continuous phase shifts | (b) 2-bit phase shifts | (c) Ground truth |
| Depth RMSE =0.148m | Depth RMSE =0.149m | |
| Depth MAE =0.1106m | Depth MAE =0.1111m | |

FIGURE 16. The 1080p depth maps for the one wall scenario at 7m separation distance are estimated for two cases of the RF phase shifters at the AR/VR device: (a) continuous phase shifts and (b) 2-bit quantized phase shifts. 16 × 16 UPA is employed with codebook oversampling factors of two. Figures (c) illustrate the ground truth depth map.

massive correlator is adopted for range estimation. Throughout this paper, two performance metrics are used: (i) root-mean-square-error (RMSE) between the estimated map and the ground truth map to indicate the standard deviation of the estimation error, and (ii) mean-absolute-error (MAE) to denote the expected value of the estimation error. The two metrics are defined in (9). Next, we evaluate the performance of our proposed mmWave MIMO depth estimation
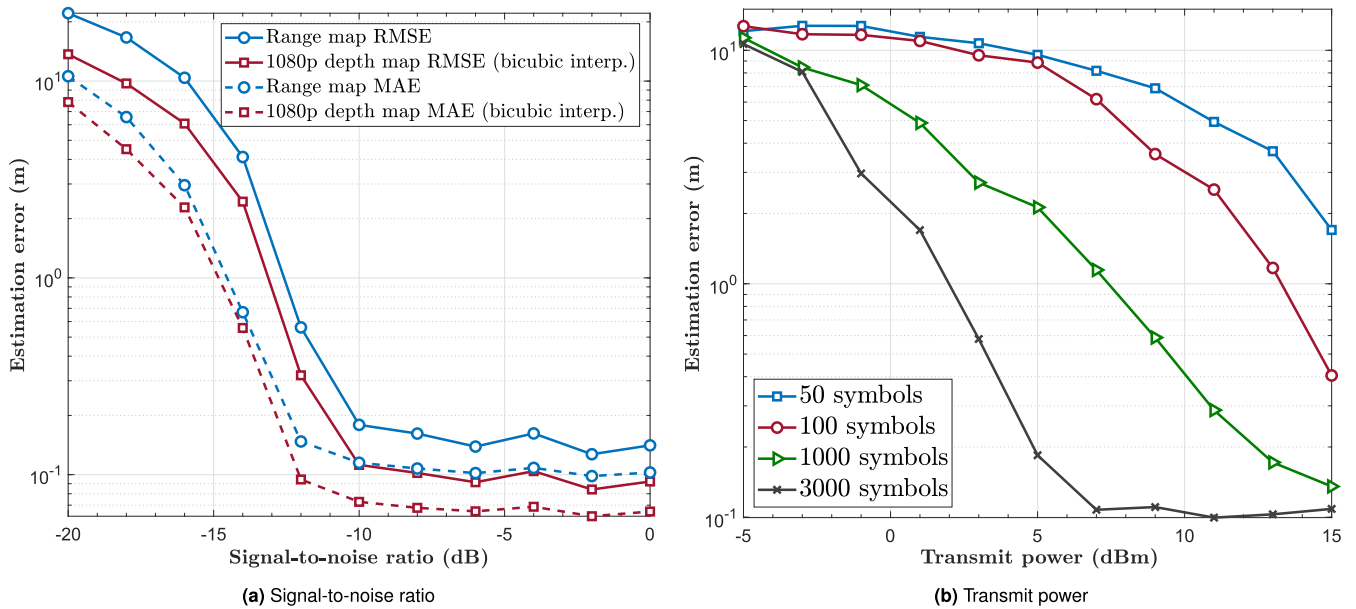
FIGURE 17. For the one wall scenario, the error performance of the proposed mmWave MIMO based depth estimation solution is evaluated under different error metrics in (a) and is evaluated for different preamble sequence lengths in (b). The wall is 7 meters away from the AR/VR device with 16 × 16 UPAs. The figures show the robustness of the developed approach under relatively low SNR regime. Note that the displayed transmit power range in (b) corresponds to average SNR range of −20.7dB to −0.7dB.

approach in four main scenarios: (i) A one wall scenario in Section VIII-B, (ii) a two walls scenario in Section VIII-C; (iii)) a room with two pillars scenario in Section VIII-D, and (iv) a conference room scenario in Section VIII-E.

B. ONE WALL SCENARIO

The one wall scenario consists of an AR/VR transceiver facing a wall in free space propagation. Unless otherwise mentioned, the separation distance between the wall and the transceiver is 7 meters and the wall building material is concrete. In Fig. 13, we show the estimated range and depth maps for the one wall scenario compared to the ground truth maps. Fig. 13(a) and Fig. 13(b) show that the range map estimation error has an average MAE of 0.098m and RMSE of 0.127m. Further, the depth map estimation error Fig. 13(c) and Fig. 13(d) has an average MAE of 0.12m and RMSE of 0.153m. Overall, these figures show that the proposed approaches can accurately estimate the range/depth maps for a wall at 7m distance from the AR/VR device with around 10cm error, which highlights the effectiveness of this approach.

Impact of the Important System Parameters: Next, we briefly evaluate the impact of the various system parameters on the performance of the proposed mmWave depth map estimation solution.

- **Number of antennas and sensing codebook beams:** In Fig. 14, we plot the estimated and ground-truth depth maps for a different number of antennas and codebook oversampling factors. As illustrated, the depth estimation accuracy can generally improve by increasing the number of antennas and/or the codebook oversampling factors. This comes with the cost of

deploying more antennas at the AR/VR device or employing more beams, which translates to a longer sensing time. In Fig. 15, we plot the estimated and ground-truth depth maps for different antenna configurations using the same number of antenna elements. As depicted, the depth estimation accuracy depends on the UPA configuration, with the best configuration being the 6 × 4 UPA because of its closeness to the 1080p aspect ratio.

- **RF phase shift quantization:** As previously described in Section VI-B, the phase quantization of the RF phase shifters in the AR/VR transceiver architecture produces a noticeable change in the radiation pattern shape of the sidelobes. To examine the effect of this phase quantization on the estimated depth maps, Fig. 16 shows the comparison of the estimated depth maps for two cases of the RF phase shifters at the AR/VR device: (a) continuous phase shift and (b) 2-bit quantized phase shifts. As depicted, the phase quantization contributes with a small negative impact on the depth map estimation accuracy for the one wall scenario at a separation distance of 7 meters.

- **Transmit sensing power:** In 17a, we investigate the effect of changing the transmit power on the depth map estimation accuracy. The SNR value of 0dB corresponds to a transmit power of 15dBm. This figure shows that a transmit power of 5dBm (SNR of −10dB) could be sufficient to reach around 10cm error for the depth estimation accuracy.

- **Preamble sequence length**: The estimation error versus transmit power is depicted in 17b for different values of preamble sequence lengths, namely preambles with 50,

TABLE 2. The estimation error results of the one wall scenario for different wall materials.

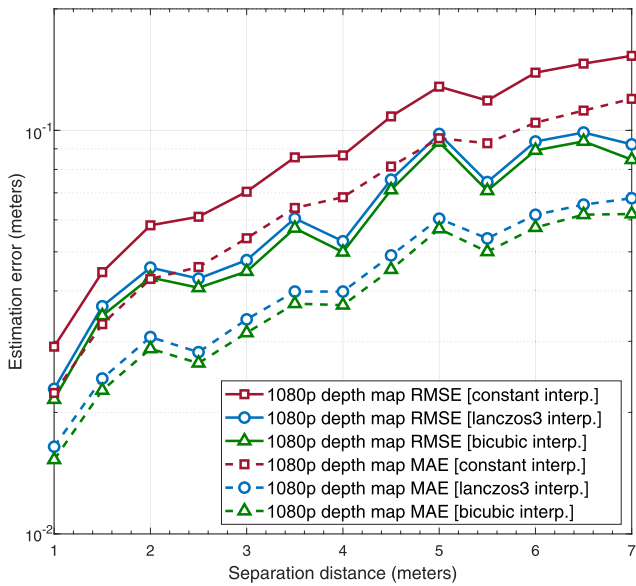| Estimation Error (meters)	Concrete	Ceilingboard	Wood	Floorboard	Layered Drywall	Glass
Basic Correlator	0.101	0.099	0.101	0.0984	0.103	12.697
Massive Correlator	0.0983	0.097	0.0983	0.0983	0.101	15.498

FIGURE 18. The error performance of the proposed mmWave MIMO based depth estimation solution is evaluated across different separation distances for the one wall scenario. The estimation error starts from ≈ 1.5m at a 1m distance and reaches around 10cm at a 7 meters distance.

(a) Two walls scenario

(a) Estimated range map

(c) Estimated depth map

(b) Ground truth range map

(d) Ground truth depth map

Range RMSE = 0.100m
Range MAE = 0.052m

Depth RMSE = 0.085m
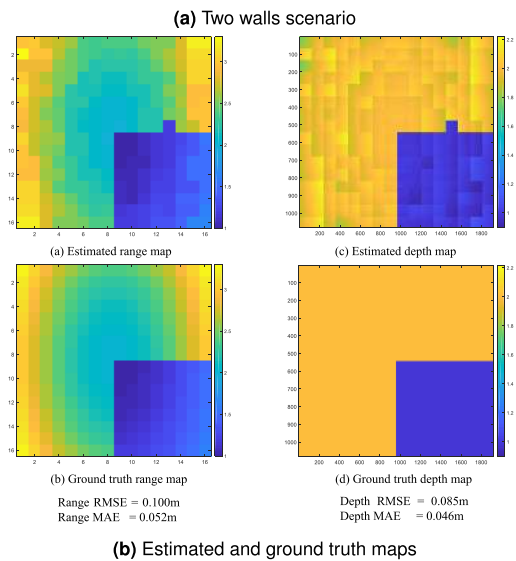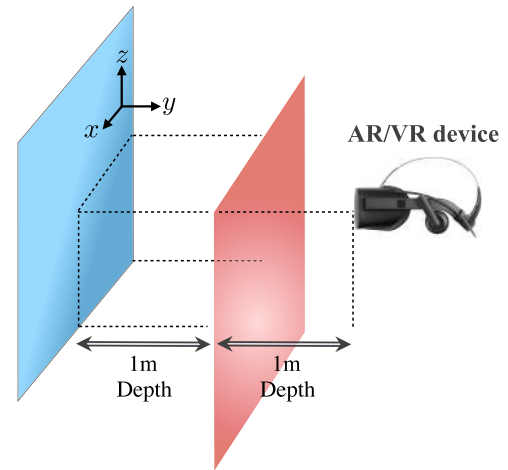Depth MAE = 0.046m

(b) Estimated and ground truth maps

FIGURE 19. (a) The adopted two walls scenario is illustrated. (b) The maps for the two walls scenario are depicted. The AR/VR device is employed with 16 × 16 UPAs. The depicted maps are the estimated maps (at the top), ground truth maps (at the bottom), range maps (on the left side), and 1080p depth maps (on the right side). Comparing (a) with (b), the range map estimation error: MAE = 0.052m. Comparing (c) with (d), the depth map estimation error: MAE = 0.046m.

100, 1000, and 3000 symbols. As shown in this figure, increasing the preamble sequence length improves the depth estimation accuracy at the expense of increased sensing time and post-processing complexity.

- **Separation distance between the AR/VR device and the facing wall:** Fig. 18 investigates the impact of increasing the depth value on the depth estimation accuracy. As shown in this figure, the larger the distance between the AR/VR device and the facing surface, the larger the error in the depth estimate, which is expected. This figure also highlights some advantage for the bicubic interpolation compared to the other interpolation methods.

- **The surface material:** Now, we evaluate the performance of the proposed approach for different surface materials. More specifically, we summarize in Table 2 the range map MAE for different candidates of the wall material. Overall, we can notice some correlation between the estimation accuracy and the *scattered to incident power ratio* property of the materials, which are summarized in Table 1.

C. TWO WALLS SCENARIO

The two walls scenario consists of one AR/VR device facing two walls in free space propagation as depicted in 19a. The

separation distance between the front wall and the AR/VR device is 1m while the separation between the back wall and the AR/VR device is 2m. The walls' building material is concrete. Each wall consists of 2, 048 faceted faces, and each face contributing with at most one backscattered ray. The purpose behind studying this scenario is to test the alignment of the estimated map compared to the ground truth depth map. The results of this test are illustrated in 19b, where the

(a) The room with two pillars
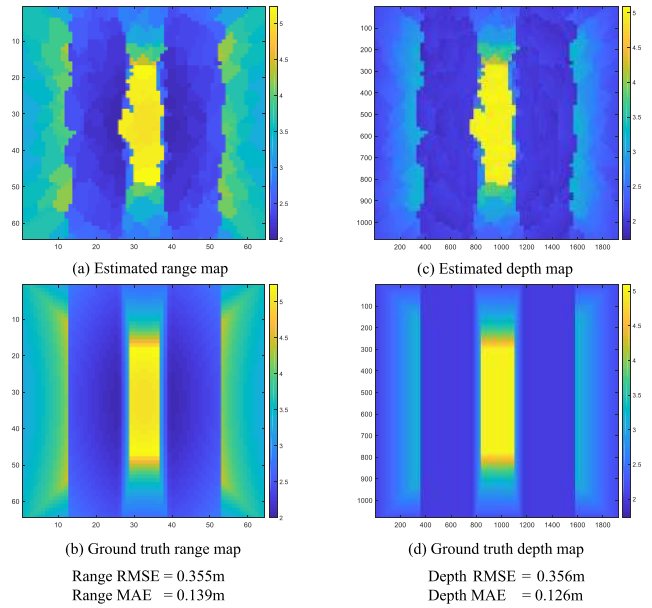

(b) The room scene from the door position
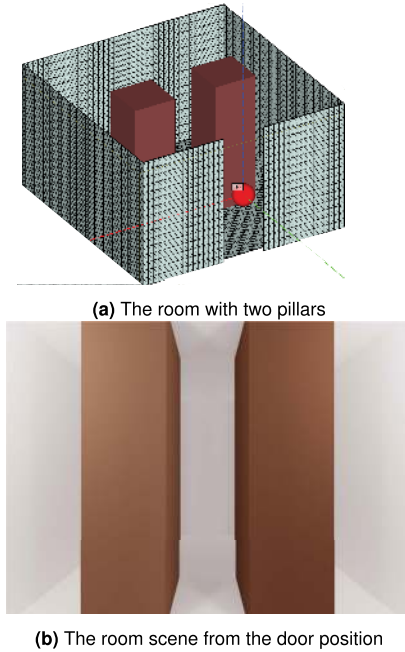
FIGURE 20. Figure (a) illustrates the bird view of the room with two pillars. Figure (b) shows the scene from the AR/VR device position, centered at the front door. The 5m×5m room consists of a concrete floor plan with two wood pillars in the middle of the room. The wood pillars are at 2 meters distance from the AR/VR device.

(a) Estimated range map

(c) Estimated depth map

(b) Ground truth range map

(d) Ground truth depth map

Range RMSE = 0.355m   Depth RMSE = 0.356m
Range MAE = 0.139m   Depth MAE   = 0.126m

FIGURE 21. The maps for the room with two pillars are depicted. 16 × 16 UPAs are employed with codebook oversampling factors of four. The depicted maps are the estimated maps (at the top), ground truth maps (at the bottom), range maps (on the left side), and 1080p depth maps (on the right side). Comparing (a) with (b), the range map estimation error: MAE = 0.139m. Comparing (c) with (d), the depth map estimation error: MAE = 0.126m.

estimated range and depth maps are compared to the ground truth maps. As shown in 19b, the two edges of the front wall in the estimated maps align reasonably well with the one displayed in the ground truth maps. This highlights the promising performance of proposed mmWave based depth estimation solution.
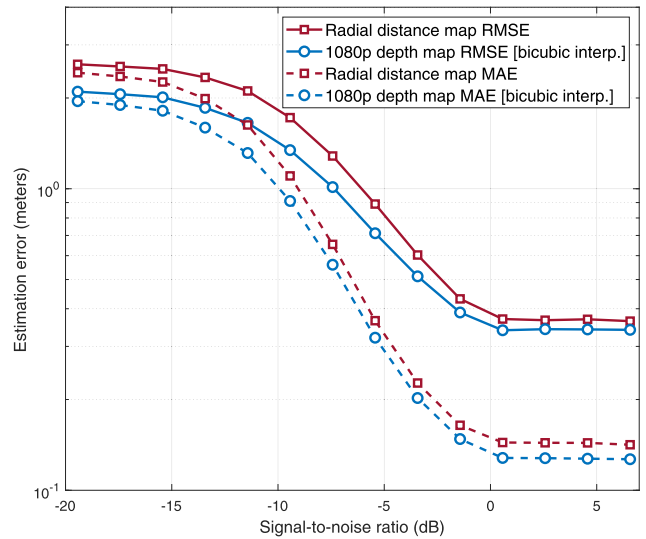
D. A ROOM WITH TWO PILLARS

In this scenario, we consider a 5m×5m room where one AR/VR device is centered at the front door of the room, as depicted in Fig. 20. The room consists of a concrete floor plan with two wood pillars in the middle of the room. The wood pillars are at 2 meters distance from the AR/VR transceiver. The floor plan consists of 15, 488 faceted faces whereas each of the wood pillars consists of 3, 072 faceted faces. Note that the ceiling of the floor plan is set to the invisible mode for visibility purposes only. For the estimation error assessment of the indoor space scenario, Fig. 21 shows the comparison between estimated and ground truth maps for 16 × 16 UPA antennas with a codebook oversampling factors of four in both azimuth and elevation dimensions.

First, Fig. 21(a) with Fig. 21(b) show the estimate and ground truth range maps, which have a MAE of 0.139m and RMSE of 0.355m. For the depth maps, Fig. 21(c) with Fig. 21(d) represent 1080p maps with estimation error of (i) 0.126m for the MAE and 0.356m for the RMSE with nearest neighbor interpolation, and (ii) 0.123m for the MAE and 0.328m for the RMSE with bicubic interpolation. From

FIGURE 22. For the room with two pillars, the error performance of the proposed mmWave MIMO based depth estimation is evaluated for different error metrics. 16 × 16 UPAs are employed with a codebook oversampling factors of four in both dimensions. This figure shows the robustness of the proposed mmWave MIMO based depth estimation under relatively low SNR regime.

observing the difference in maps, the mmWave reasonably recover most of the depth information of the scene with low codebook resolution (16 × 16) compared to the ground truth 1080p resolution. With narrower transmit and receive beams, i.e. more antenna elements, the estimation accuracy is expected to further improve.

(b) First scene

(a) Conference room scenario

(c) Second scene

FIGURE 23. (a) the bird view of the conference room scenario; (b) and (c) the scenes under study. The 10m×10m indoor space contain a 6m×6m conference room in glass. the indoor space walls are made from layered drywall, the ceiling is made from ceiling board and the floor is made from floorboard. The conference room chairs and tables are made from wood.

The depth map estimation accuracy for this scenario is also evaluated at different SNRs in Fig. 22. In this figure, we adopt the model and system parameters used in Fig. 21 with 16×16 UPAs and oversampling factors of four. It is also worth mentioning that 0dB SNR corresponds to −20dBm transmit power in our setup. As shown in Fig. 22, the estimated depth maps have MAE of almost 10cm at 0dB, which highlights the promising performance of our proposed depth map estimation approach at relatively low SNRs and in an indoor room with several surfaces and different materials. This will be further emphasized in the following subsection.
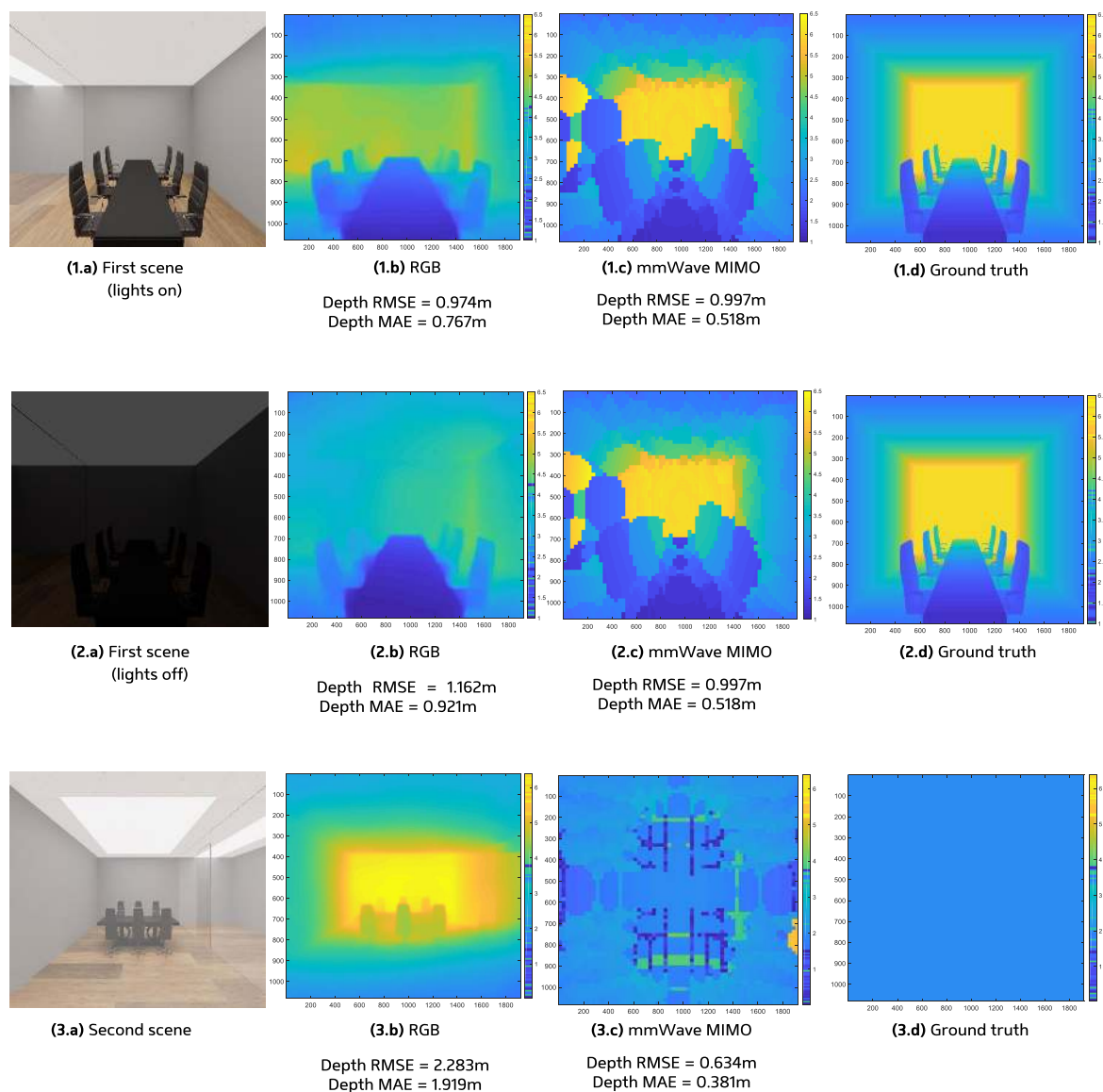
E. CONFERENCE ROOM SCENARIO

In this scenario, we consider the conference room shown in Fig. 23. The ceiling of the indoor space is set to the invisible mode for visibility purpose only. The 10m×10m indoor space has a 6m×6m conference room with glass walls. The indoor space walls are made of layered drywall, the ceiling is made of ceiling board, and the floor is made of floorboard. The conference room chairs and tables are made of wood. The conference room door opening is 1m in width and 2.7m in height. The number of facets for each item in the indoor space is as follows: $2,048$ facets for the layered drywall, $2,048$ facets for the floorboard, and $2,048$ facets for the ceiling board. In addition, the number of facets for each item in the conference room is as follows: $1,568$ facets for the glass wall, $4,446$ facets for the table, $21,192$ facets for the office chairs. The conference room scenario consists of two AR/VR devices for two scenes under study — the first device is centered at the front door of the conference room while

the second transceiver is placed outside of the conference room facing the other glass facet. The scenes captured by the AR/VR camera for the two cases are shown in Fig. 23(b) and Fig. 23(c).

One main motivation for leveraging mmWave MIMO to estimate the depth maps (compared to RGB based depth estimation approaches) is the expected higher efficiency in detecting transparent and dark objects. In Fig. 24, we compare our mmWave MIMO based depth estimation approach with the RGB based depth estimation approach, detailed in [2], for the two considered conference room scenarios. It's worth emphasizing here that the algorithms in [2] achieve considerably good depth accuracy when tested on the NYU depth V2 dataset [45]. As shown in Fig. 24, the mmWave MIMO based estimator outperforms the RGB based estimator in recognizing transparent and dark objects. For the first scene, the glass wall was not detected by the RGB estimator. Also, in the presence of a scene with low illumination, the mmWave MIMO based estimator performance shows robustness in the estimation accuracy compared to the RGB based estimator. Figure 1.c) and 2.c) were generated with the aid of the SLR approach in Section VI-B, with $\delta_H = 2, \delta_V = 3$. For this reason, the depth maps constructed by the mmWave MIMO system seem coarser than the one constructed by RGB cameras, which can be resolved using morphological image processing operations, e.g., the erosion operation. As for the second scene, the RGB based estimator is unable to detect the transparent glass compared to the mmWave MIMO based estimator. Interestingly, despite the fact that the glass scattering ratio is 0% based on Table 1, the conference room

(1.a) First scene
(lights on)

(1.b) RGB

Depth RMSE = 0.974m
Depth MAE = 0.767m

(1.c) mmWave MIMO

Depth RMSE = 0.997m
Depth MAE = 0.518m

(1.d) Ground truth

(2.a) First scene
(lights off)

(2.b) RGB

Depth RMSE = 1.162m
Depth MAE = 0.921m

(2.c) mmWave MIMO

Depth RMSE = 0.997m
Depth MAE = 0.518m

(2.d) Ground truth

(3.a) Second scene

(3.b) RGB

Depth RMSE = 2.283m
Depth MAE = 1.919m

(3.c) mmWave MIMO

Depth RMSE = 0.634m
Depth MAE = 0.381m

(3.d) Ground truth

FIGURE 24. For the conference room scenario, the proposed mmWave MIMO based depth estimation is compared with the RGB based depth estimation in [2]. 16×16 UPAs are employed with codebook oversampling factors of four. The depicted maps are the maps of the first scene with lights on/off (The top two rows) and the second scene (the bottom row). (a) The scenes under study; (b) the estimated maps from monocular RGB images; (c) the estimated maps from our proposed solution; (d) the ground truth depth maps.

glass wall is partially recovered by the mmWave MIMO based estimator because of the boresight reflection path. This makes the wireless AR/VR experience safer by providing the ability to detect transparent surfaces. All these promising results highlight the potential of leveraging the proposed mmWave MIMO based depth map estimation approaches for immersive AR/VR experience.

IX. CONCLUSION

In this paper, we considered the problem of estimating accurate depth maps for AR/VR devices, which is an essential goal for immersive mixed-reality experience. For this problem, we proposed leveraging the mmWave communication systems that are deployed on the AR/VR devices to estimate

and build high-resolution depth maps. We formulated the communication-constrained depth map sensing problem and proposed a comprehensive framework for realizing this objective. The proposed framework includes (i) the construction of depth map specific sensing codebooks using practical mmWave antenna arrays and (ii) the development of efficient post-processing solutions for jointly processing the receive signals from the multiple sensing beams and estimating high-resolution depth maps. Simulations using accurate 3D ray-tracing models confirmed the promising accuracy of our proposed mmWave based depth map estimation approach in various environment scenarios. In particular, the results show that the proposed approach can construct relatively high-resolution depth maps with less than 10cm error using

practical mmWave systems. This highlights the potential of leveraging this solution to complement RGB-D based depth maps and realize immersive depth perception for wireless virtual/augmented reality systems.

REFERENCES

[1] VIVE. *VIVE Wireless Adapter*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.vive.com/us/accessory/wireless-adapter/

[2] J. Hu, M. Ozay, Y. Zhang, and T. Okatani, "Revisiting single image depth estimation: Toward higher resolution maps with accurate object boundaries," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1043–1051.

[3] F. Ma and S. Karaman, "Sparse-to-dense: Depth prediction from sparse depth samples and a single image," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2018, pp. 4796–4803.

[4] G. Riegler, Y. Liao, S. Donne, V. Koltun, and A. Geiger, "Connecting the dots: Learning representations for active monocular depth estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 7624–7633.

[5] Y. Zhang, S. Khamis, C. Rhemann, J. Valentin, A. Kowdle, V. Tankovich, M. Schoenberg, S. Izadi, T. Funkhouser, and S. Fanello, "Activestereonet: End-to-end self-supervised learning for active stereo systems," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 784–801.

[6] C. Cheng, Y. Wang, C. Wei, C. Chen, and L. Lin, "IR pattern characteristics for active stereo matching," U.S. Patent 16 359 699, Oct. 3, 2019.

[7] T. Gruber, M. Kokhova, W. Ritter, N. Haala, and K. Dictmayer, "Learning super-resolved depth from active gated imaging," in *Proc. 21st Int. Conf. Intell. Transp. Syst. (ITSC)*, Nov. 2018, pp. 3051–3058.

[8] T. Gruber, F. Julca-Aguilar, M. Bijelic, and F. Heide, "Gated2Depth: Real-time dense lidar from gated images," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2019, pp. 1506–1516.

[9] Intel. *Intel RealSense Depth Camera D435*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.intelrealsense.com/depth-camera-d435/

[10] Y. Wang, W.-L. Chao, D. Garg, B. Hariharan, M. Campbell, and K. Q. Weinberger, "Pseudo-LiDAR from visual depth estimation: Bridging the gap in 3D object detection for autonomous driving," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 8445–8453.

[11] Y. Zhang and T. Funkhouser, "Deep depth completion of a single RGB-D image," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 175–185.

[12] J. A. Zhang, A. Cantoni, X. Huang, Y. J. Guo, and R. W. Heath, Jr., "Joint communications and sensing using two steerable analog antenna arrays," in *Proc. IEEE 85th Veh. Technol. Conf. (VTC Spring)*, Jun. 2017, pp. 1–5.

[13] P. Kumari, J. Choi, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "IEEE 802.11ad-based radar: An approach to joint vehicular communication-radar system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 4, pp. 3012–3027, Apr. 2018.

[14] J. Lien, N. Gillian, M. E. Karagozler, P. Amihood, C. Schwesig, E. Olson, H. Raja, and I. Poupyrev, "Soli: Ubiquitous gesture sensing with millimeter wave radar," *ACM Trans. Graph.*, vol. 35, no. 4, pp. 1–19, Jul. 2016.

[15] T. Wei and X. Zhang, "MTrack: high-precision passive tracking using millimeter wave radios," in *Proc. 21st Annu. Int. Conf. Mobile Comput. Netw.*, Sep. 2015, pp. 117–129.

[16] Remcom. *Wireless Insite*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.remcom.com/wireless-insite-em-propagation-software

[17] *Blender 2.80*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.blender.org/

[18] *Blenderkit*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.blenderkit.com/

[19] *Blender Demo Files*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.blender.org/download/demo-files/

[20] *Turbosquid*. Accessed: Mar. 23, 2021. [Online]. Available: https://www.turbosquid.com/Search/3D-Models

[21] *Free3d*. Accessed: Mar. 23, 2021. [Online]. Available: https://free3d.com/

[22] A. Sabharwal, P. Schniter, D. Guo, D. W. Bliss, S. Rangarajan, and R. Wichman, "In-band full-duplex wireless: Challenges and opportunities," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 9, pp. 1637–1652, Sep. 2014.

[23] N. A. Estep, D. L. Sounas, J. Soric, and A. Alù, "Magnetic-free non-reciprocity and isolation based on parametrically modulated coupled-resonator loops," *Nature Phys.*, vol. 10, no. 12, pp. 923–927, Dec. 2014.

[24] T. Dinc and H. Krishnaswamy, "A 28 GHz magnetic-free non-reciprocal passive CMOS circulator based on spatio-temporal conductance modulation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 294–295.

[25] J. Zhou, N. Reiskarimian, and H. Krishnaswamy, "Receiver with integrated magnetic-free N-path-filter-based non-reciprocal circulator and baseband self-interference cancellation for full-duplex wireless," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Jan. 2016, pp. 178–180.

[26] A. Nagulu and H. Krishnaswamy, "Non-magnetic 60 GHz SOI CMOS circulator based on Loss/Dispersion-engineered switched bandpass filters," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2019, pp. 446–448.

[27] R. W. Heath, Jr., N. González-Prelcic, S. Rangan, W. Roh, and A. M. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Apr. 2016.

[28] A. Alkhateeb, J. Mo, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "MIMO precoding and combining solutions for millimeter-wave systems," *IEEE Commun. Mag.*, vol. 52, no. 12, pp. 122–131, Dec. 2014.

[29] M. A. Richards, J. A. Scheer, and W. A. Holm, *Principles of Modern Radar: Basic Principles*. Raleigh, NC, USA: SciTech, 2010.

[30] P. Kumari, S. A. Vorobyov, and R. W. Heath, Jr., "Adaptive virtual waveform design for millimeter-wave joint communication-radar," 2019, *arXiv:1904.05516*. [Online]. Available: http://arxiv.org/abs/1904.05516

[31] Q. H. Spencer, B. D. Jeffs, M. A. Jensen, and A. L. Swindlehurst, "Modeling the statistical time and angle of arrival characteristics of an indoor multipath channel," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 347–360, Mar. 2000.

[32] P. Smulders, "Statistical characterization of 60-GHz indoor radio channels," *IEEE Trans. Antennas Propag.*, vol. 57, no. 10, pp. 2820–2829, Oct. 2009.

[33] O. E. Ayach, S. Rajagopal, S. Abu-Surra, Z. Pi, and R. W. Heath, Jr., "Spatially sparse precoding in millimeter wave MIMO systems," *IEEE Trans. Wireless Commun.*, vol. 13, no. 3, pp. 1499–1513, Mar. 2014.

[34] A. Guerra, F. Guidi, D. Dardari, A. Clemente, and R. D'Errico, "A millimeter-wave indoor backscattering channel model for environment mapping," *IEEE Trans. Antennas Propag.*, vol. 65, no. 9, pp. 4935–4940, Sep. 2017.

[35] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.

[36] P. Bidigare, U. Madhow, R. Mudumbai, and D. Scherber, "Attaining fundamental bounds on timing synchronization," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 5229–5232.

[37] A. Herschfelt, H. Yu, S. Wu, H. Lee, and D. W. Bliss, "Joint positioning-communications system design: Leveraging phase-accurate time-of-flight estimation and distributed coherence," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 433–437.

[38] A. Herschfelt, "Simultaneous positioning and communications: Hybrid radio architecture, estimation techniques, and experimental validation," Ph.D. dissertation, School Elect., Comput. Energy Eng., Arizona State Univ., Tempe, AZ, USA, 2019.

[39] A. Alkhateeb and R. W. Heath, Jr., "Frequency selective hybrid precoding for limited feedback millimeter wave systems," *IEEE Trans. Commun.*, vol. 64, no. 5, pp. 1801–1818, May 2016.

[40] W. L. Melvin and J. A. Scheer, *Principles of Modern Radar Volume II: Advanced Techniques*. Edison, NJ, USA: SciTech, 2013.

[41] M. I. Dessouky, H. A. Sharshar, and Y. A. Albagory, "Efficient sidelobe reduction technique for small-sized concentric circular arrays," *Prog. Electromagn. Res.*, vol. 65, pp. 187–200, Nov. 2006.

[42] Y. A. Albagory, M. Dessouky, and H. Sharshar, "An approach for low sidelobe beamforming in uniform concentric circular arrays," *Wireless Pers. Commun.*, vol. 43, no. 4, pp. 1363–1368, Nov. 2007.

[43] E. Grossi, M. Lops, and L. Venturino, "Detection and localization of multiple targets in IEEE 802.11ad networks," in *Proc. 53rd Asilomar Conf. Signals, Syst., Comput.*, Nov. 2019, pp. 947–951.

[44] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. Inf. Theory Appl. Workshop (ITA)*, San Diego, CA, USA, Feb. 2019, pp. 1–8.

[45] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2012, pp. 746–760.

ABDELRAHMAN TAHA received the B.Sc. (Hons.) and M.Sc. degrees in electrical engineering from Ain Shams University, Cairo, Egypt, in 2013 and 2018, respectively. He is currently pursuing the Ph.D. degree with the School of Electrical, Computer and Energy Engineering, Arizona State University (ASU), with a specialization in signal processing systems and communications. His research interests include massive MIMO systems and signal processing for wireless communications. He received the Egyptian Government Excellence Award in 2013, for being ranked first among all engineering programs, B.Sc. valedictorian, Faculty of Engineering, Ain Shams University.

QI QU received the Ph.D. degree from UC San Diego, in 2008. He is currently with the Facebook Reality Laboratory leading the wireless architecture development. Prior to that, he was with the Facebook Connectivity Laboratory focusing on Massive MIMO and other advanced technologies with the aim of connecting un(under)-connected population, and investigation of machine-learning and deep-learning in wireless mmwave communications. Before joining Facebook, he was the Project Lead with Broadcom driving the 60GHz baseband chipset design for one of Broadcom's largest customers. Previously, he was with several other companies (Intel, Maxim, etc.,) and worked on various connectivity technologies, such as LTE-A, Wi-Fi, and powerline OFDM modem. He has also participated as a major contributor for the development of IEEE/ITU International standards. His current research interests include advanced wireless communications technologies, AR/VR, and machine learning applications.

SAM ALEX received the B.Tech. degree from the Cochin University of Science and Technology, India, and the M.Tech. degree from IIT Madras, India. From 2004 to 2015, he has held engineering positions with Broadcom Inc., Quantenna Communications, where he was involved in designing and architecting the modem physical layer for various standards (WiMax, LTE, and WiFi 11ac/ax). Since 2015, he has been with Facebook Inc., currently playing the role of a Systems Architect. His current research interests include communication systems, AR/VR systems, machine learning, and brain inspired AI.

PING WANG received the B.Eng. and M.Eng. degrees in electrical engineering from the Beijing University of Posts and Telecommunications, and the Ph.D. degree from McGill University, Canada. Before joining Facebook, she worked with Apple on wireless software organization. Prior with Apple, she worked as a Staff Research Scientist with the Wireless Communication Laboratory under Intel Labs, where her research was focused on system design in advanced cellular and WiFi wireless networks. She also worked as a Principal Member of Technical Staff on radio access networks of AT&T Labs and a System Engineer with Ericsson, Canada, on LTE modem development. She is currently a Wireless System Architect with the Facebook's Reality Laboratory (FRL).

WILLIAM L. ABBOTT received the B.S. degree (Hons.) in electrical engineering from Princeton University, the M.S. degree in electrical engineering from the University of California, Berkeley, and the Ph.D. degree in electrical engineering from Stanford University. Prior to joining Facebook, he was in the roles of a Senior Director of Technology at Qualcomm, Inc., developing WiFi6 chipsets, and a Vice President of Product Development at ASSIA, Inc., developing broadband network optimization systems. He is currently leading the FRL Wireless Systems Engineering Group at Facebook, Inc.

AHMED ALKHATEEB received the B.S. degree (Hons.) and the M.S. degree in electrical engineering from Cairo University, Egypt, in 2008 and 2012, respectively, and the Ph.D. degree in electrical engineering from The University of Texas at Austin, USA, in August 2016. From September 2016 to December 2017, he was a Wireless Communications Researcher with the Connectivity Laboratory, Facebook, Menlo Park, CA, USA. He joined Arizona State University (ASU) in Spring 2018, where he is currently an Assistant Professor with the School of Electrical, Computer and Energy Engineering. He has held Research and Development internships with FutureWei Technologies (Huawei), Chicago, IL, USA, and Samsung Research America (SRA), Dallas, TX, USA. His research interests include wireless communications, communication theory, signal processing, machine learning, and applied mathematics. He was a recipient of the 2012 MCD Fellowship from The University of Texas at Austin and the 2016 IEEE Signal Processing Society Young Author Best Paper Award for his work on hybrid precoding and channel estimation in millimeter wave communication systems.

● ● ●