

Rascene: High-Fidelity 3D Scene Imaging with mmWave Communication Signals

Supplementary Material

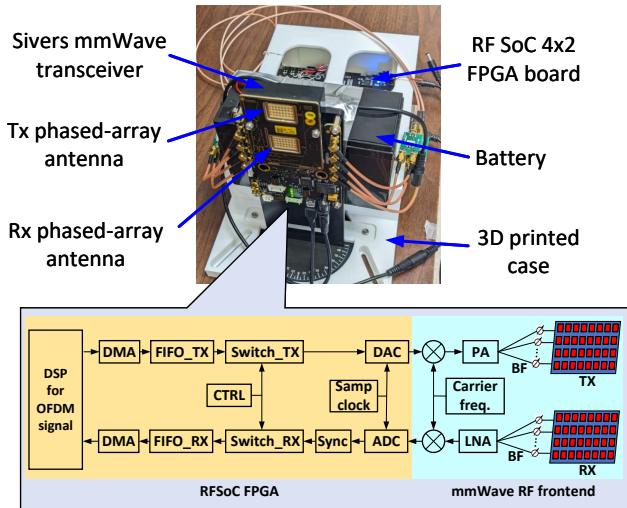


Figure 13. Our prototyped monostatic ISAC device.

8. Monostatic ISAC Hardware

We built a monostatic ISAC prototype using commercial off-the-shelf (COTS) components, enabling joint communication and sensing within a compact device.

Fig. 13 shows our monostatic ISAC prototype, with its parameters summarized in Tab. 7. The system consists of two primary COTS modules: (i) an AMD/Xilinx RFSoc 4x2 FPGA board, and (ii) a Sivers mmWave transceiver with Tx/Rx phased-array antennas. The RFSoc FPGA implements an OFDM signal processing pipeline compatible with 5G and Wi-Fi protocols, while the mmWave transceiver handles 60 GHz radio transmission and reception. Within the FPGA, the transmission and reception pipelines are jointly optimized and precisely calibrated to ensure timing and phase alignment required for monostatic sensing. Phased-array antenna control is seamlessly integrated into the processing pipeline, enabling beam steering to be synchronized with signal transmission.

Fig. 14 shows an example of simultaneous sensing and communication using the prototyped Rascene. During sensing data collection, Rascene simultaneously sends data packets to another device, supporting continuous video streaming. Both sensing and communication share the same hardware, spectrum band, and radiated energy. Our implementation demonstrates that a monostatic ISAC device can be realized without specialized sensing hardware. Furthermore, because the design leverages COTS communication components, it can be readily integrated into existing 5G and Wi-Fi mmWave communication devices through firmware and software upgrades.

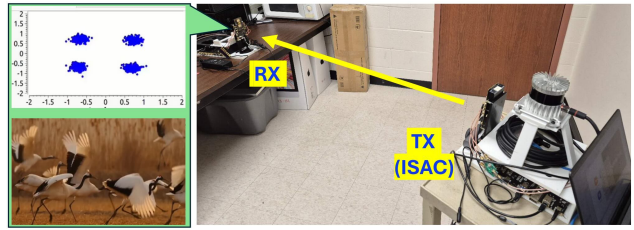


Figure 14. Illustration of video streaming data communication during sensing data collection.

Table 7. The parameters of our monostatic ISAC device.

Hardware parameters	
Sampling rate	1.2288 GSPS
Number of Tx antennas	16
Number of Rx antennas	16
Center frequency	60 GHz
Transmission power	20 dBm
Communication parameters	
Waveform	OFDM
FFT points	1024
Number of valid subcarriers	900
Cyclic prefix length	276
OFDM symbol duration	1.057 μ s
Number of OFDM symbols per frame	$16 \times 16 = 256$
Supporting protocols	5G and Wi-Fi
Sensing parameters	
Detection time of a frame	67 μ s
Number of frames per second	10
Theoretical detection range	30 m
Practical detection range	10 m
Number of effective antennas (horizontal)	8
Number of effective antennas (elevation)	4
Horizontal antenna spacing	0.5 wavelength
Elevation antenna spacing	0.5 wavelength
Horizontal field of view	$[-60^\circ, 60^\circ]$
Elevation field of view	$[-30^\circ, 30^\circ]$

9. Data Collection

Platform. To collect paired RF-LiDAR data, we mounted our custom-designed ISAC device, an Ouster OS0-128 LiDAR, and a TDK ICM-20948 IMU on a movable cart. The final dataset contains synchronized RF-LiDAR frame pairs collected from 20 indoor environments spanning diverse layouts, clutter levels, and construction materials such as drywall, glass, and metal. For each environment, we recorded approximately 10-20 minutes of data while manually moving the cart along unconstrained trajectories. Ex-

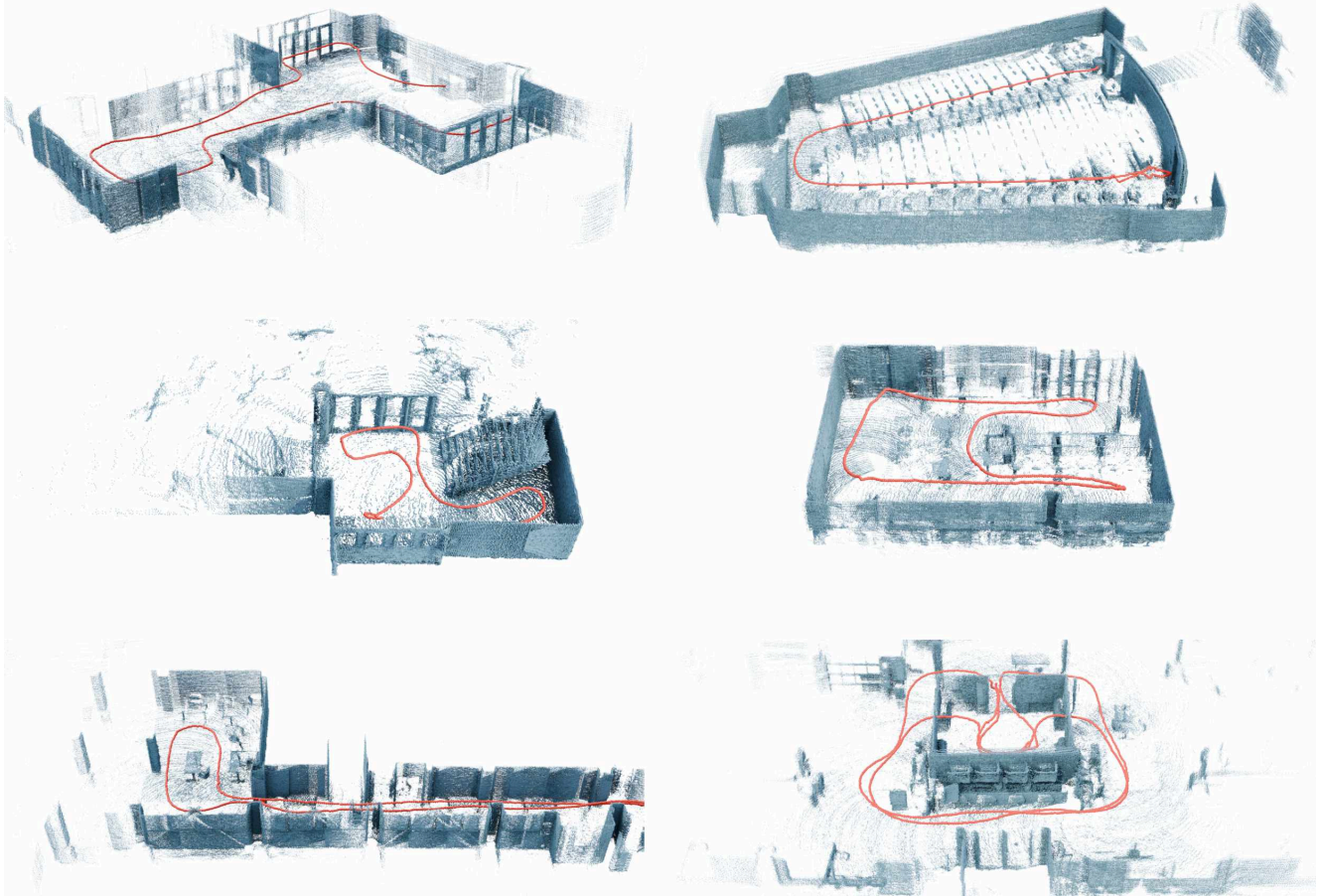


Figure 15. Sample trajectory segments (shown in red) from different scenes, visualized on the ground truth LiDAR point clouds.

ample trajectory segments and scene snapshots are shown in Fig. 15 and Fig. 16, respectively. All sensing modalities, including radio frames, LiDAR scans, and IMU measurements, were synchronized using timestamps from a shared clock source.

Ground Truth. The Ouster OS0-128 LiDAR provides a 360° horizontal and 90° vertical field of view, whereas our ISAC sensor covers 120° horizontally and 60° vertically. To align the two modalities, we calibrated the fixed extrinsic transformation between the rigidly mounted sensors and used it to crop the panoramic LiDAR observations to the field of view of the ISAC sensor. The cropped high-resolution LiDAR point clouds are then used to derive the ground-truth 3D geometry V^* and depth maps D^* for training and evaluation.

Temporal Sampling Strategy. Both the LiDAR and ISAC streams are recorded at 10 Hz, while the platform moves at an average speed of 0.5 m/s. Directly using consecutive frames would yield only a small spatial baseline and limited parallax, making it difficult to disambiguate true scene structure from multipath artifacts. We therefore adopt a sparse temporal sampling strategy: one frame is selected

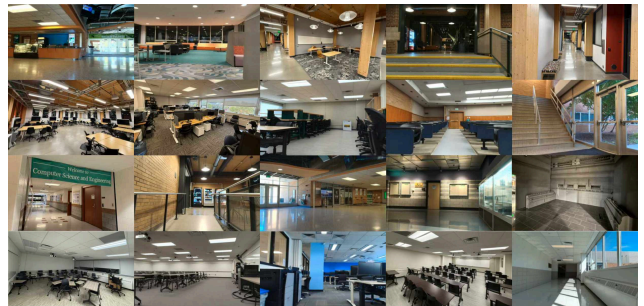


Figure 16. Example snapshots from the 20 distinct indoor environments included in our dataset.

every 2 s from the continuous streams, and five sampled frames are grouped into one input window. This design preserves sufficient spatial overlap for cross-frame geometric consensus while introducing enough viewpoint variation to provide useful parallax and more diverse multipath observations. After warping the frames into a shared reference coordinate system, true scene structures remain more consistent across views than multipath artifacts, which makes the fusion process more reliable.