High-resolution mmWave Imaging using Metasurface and Diffusion

Yida Wang[1†], Yu Lu[1†], Yuxuan Zhou[2†], Yifei Shen[3], Lili Qiu[3,4], Zeyuan Lai[5†], Yi-Chao Chen[1*], Hao Pan[3], Juntao Zhou[1], Dian Ding[1], Mei Wang[4], Guangtao Xue[1], Qian Zhang[2]
[1]Shanghai Jiao Tong University, China, [2]The Hong Kong University of Science and Technology, China, [3]Microsoft Research Aisa, China, [4]UT Austin, USA, [5]University of Science and Technology of China, China
{yidawang,yulu01,yichao,juntaozhou,dingdian94,gt_xue}@sjtu.edu.cn,{yzhoudo,yshenaw}@connect.ust.hk,
{liliqiu,panhao}@microsoft.com,lzy1522659928@mail.ustc.edu.cn,meiwang@utexas.edu,qianzh@ust.hk

Abstract

In this paper, we propose a novel high-resolution mmWave imaging technique that operates with a small, off-the-shelf mmWave module and eliminates the need for any mechanical movement, offering a streamlined, portable solution. Our approach tackles two primary challenges: 1) mmWave commodity hardware is constrained by a limited number of antennas, limiting imaging resolution, and 2) most wireless imaging algorithms rely on compressive sensing to overcome the physical constraints, which assumes sparsity – a condition that may not always apply. To address these challenges, we first design an optimized mmWave metasurface specifically tailored for high-resolution imaging. This involves deriving a unit cell pattern that achieves high signal penetration and near-2π phase control, followed by joint optimization of both the metasurface and the codebook to further refine the signal quality and imaging resolution. We further propose a diffusion-based neural network model that transforms mmWave signals into high-quality images by directly exploiting the inherent features of target images, providing a robust alternative to conventional compressive sensing approaches. Our method encodes mmWave signals into physical representations and employs conditional generation through stable diffusion, effectively enhancing image quality. Through comprehensive implementation and rigorous testbed experiments, we demonstrate the feasibility and effectiveness of our approach.

CCS Concepts

- **Human-centered computing → Ubiquitous and mobile computing systems and tools**.

Keywords

mmWave Imaging, mmWave Metasurface, Diffusion Model

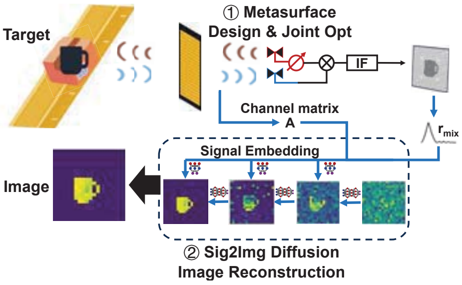


Figure 1: Illustration of MIMSID.

ACM Reference Format:
Yida Wang[1†], Yu Lu[1†], Yuxuan Zhou[2†], Yifei Shen[3], Lili Qiu[3,4], Zeyuan Lai[5†], Yi-Chao Chen[1*], Hao Pan[3], Juntao Zhou[1], Dian Ding[1], Mei Wang[4], Guangtao Xue[1], Qian Zhang[2]. 2025. High-resolution mmWave Imaging using Metasurface and Diffusion. In *The 23rd Annual International Conference on Mobile Systems, Applications and Services (MobiSys '25), June 23–27, 2025, Anaheim, CA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3711875.3729162

1 Introduction

Motivation: With the fast development of commercial off-the-shelf (COTS) mmWave radars and the rising need for fine-grain imaging in obstructed or privacy-preserving conditions, mmWave imaging has become an attractive solution.

Traditional mmWave imaging uses beamforming in both azimuth and elevation dimensions to retrieve reflected signals in each direction. However, accurate beamforming requires a large antenna array in both azimuth and elevation dimensions to achieve high angle resolutions (below 1 degree), which is not affordable for COTS mmWave radars (*e.g.*, MMWCAS-RF-EVM from Texas Instruments with 12 transmitters and 16 receivers can only achieve 1.4 degrees and 18 degrees for azimuth and elevation angle resolution, respectively [55]).

To reduce the size of the antenna array, many existing mmWave imaging methods apply Synthetic Array Radar (SAR) [14, 24, 30, 40, 43, 45, 60] to extend the virtual aperture of mmWave transceiver arrays for high-resolution imaging. However, SAR-based methods require guide rails or other scanning approaches to perform mechanical scanning, resulting in large machinery sizes and potential risks with movements, such as collisions. [24] applies a rotation SAR with a compacted size, but still requires machinery rotations. [43] achieves high-resolution imaging with handheld devices, but requires human control over handheld devices. SAR-based approaches are hard to deploy in compacted spaces (*e.g.*, rotary assembly lines [41]) or unmanned inspection tasks, forcing us to seek a stationary mmWave imaging solution with a compacted size.

Computational imaging is an interdisciplinary field that combines imaging hardware with computer algorithms to create novel imaging systems capable of generating single-shot high-resolution images.

$*$ Corresponding author. † Complete this work as interns at Microsoft Research Asia.

MobiSys '25, Anaheim, CA, USA

ACM ISBN 979-8-4007-1453-5/2025/06
https://doi.org/10.1145/3711875.3729162

Compressive sensing is a common technique in computational imaging that captures and reconstructs high-dimensional data with fewer measurements. While this technique has been widely used to turn wireless signals into images under limited measurements and insufficient antenna arrays [12], it faces several critical limitations: 1) regarding formulation, it assumes a special structure about images (*e.g.*, sparse or low-rank). However, these constraints cannot fully characterize the distribution of real images (*e.g.*, not all images exhibit these properties, and there may be additional properties in the real images that are not captured by the sparsity); 2) Existing algorithms are either optimization-based (*e.g.*, ADMM [2]) or message-passing-based (*e.g.*, AMP [10]). Optimization-based algorithms may suffer from sub-optimality and require many iterations to converge, while message-passing-based algorithms require strict assumptions on sensing matrices and images [1].

Our approach: In this paper, we leverage a stationary COTS mmWave radar to perform sub-centimeter resolution imaging, which faces several challenges: 1) The insufficient antenna array limits the resolution in both the azimuth and elevation direction for sub-centimeter imaging. 2) Real-world images do not always satisfy the sparse or low-rank assumption required by compressive sensing methods. Therefore, we propose MIMSID to address the above challenges by 1) designing a passive mmWave metasurface to optimize the measurement matrix $\boldsymbol{A}$ and increase imaging resolution, and 2) developing a conditional diffusion network to reconstruct accurate images based on target reflection signals.

Expanding phased array: The imaging resolution depends on the measurement matrix $\boldsymbol{A}$, which is dictated by the hardware setup (*i.e.*, the number of TXs and RXs). However, the cost of mmWave devices increases rapidly with the number of antennas. In addition, the largest number of antennas is up to 12 TXs × 16 RXs for a COTS radar, still insufficient for sub-centimeter-level imaging.

Motivated by recent works on expanding phased arrays with metasurfaces [12, 13, 35], we design a metasurface-based mmWave imaging system. We use a passive metasurface for its low cost and ease of deployment. Our design achieves two important properties: 1) low signal attenuation with wide phase control in a high-frequency band (77-81 GHz), and 2) jointly optimized passive metasurface phase map and transmitter codebooks for optimal imaging performance.

Signal-to-image diffusion: To address the limitations of compressive sensing, we design a *Sig2Img Diffusion* for accurate image generation based on several theoretical observations: 1) The power of diffusion models lies in their unique ability to learn the characteristics of real images and incorporate these characteristics into the generation process [48]. They are promising to address and overcome the restricted sparsity assumption. 2) In Sec. 5.2, we establish the connections between diffusion models and classic message-passing algorithms, demonstrating their optimality in signal recovery. Unlike message-passing algorithms, diffusion models automatically learn the image distribution rather than imposing sparsity or low-rank assumption, which may not hold in general, thereby further improving performance. Our conditional diffusion model takes the measurement matrix $\boldsymbol{A}$ and the target signal $\boldsymbol{r}$ as input and outputs the resulting image.

Channel estimation: In addition to designing a metasurface-based mmWave system and an effective inference algorithm, we need to get the up-to-date measurement matrix $\boldsymbol{A}$ associated with $\boldsymbol{r}$ since our system takes paired $\boldsymbol{A}$ and $\boldsymbol{r}$ as inputs. We develop a compressive sensing algorithm to estimate $\boldsymbol{A}$ using a few known template images (*i.e.*, solving $\boldsymbol{A}$ based on the known images $\boldsymbol{x}$ and its corresponding signals $\boldsymbol{r}$ where $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x}$). In this way, we can efficiently adapt to the changes in the mmWave channel.

System implementation: We implement our system, MIMSID, as shown in Fig. 1. It first jointly optimizes the passive metasurface and codewords at the transmitters. Then we deploy the metasurface along with the mmWave module in the testbed. We train the signal-to-image diffusion model to map wireless signals to images using the training data generated from the simulator based on the measurement matrix $\boldsymbol{A}$s. Then we apply the diffusion model to infer an image based on the measured reflection signal of the target and its corresponding measurement matrix collected from the testbed.

Contribution: Our contributions are as follows:

- We design a novel passive transmissive mmWave metasurface that significantly enhances imaging performance by jointly optimizing transmitter (TX) codewords and the metasurface's phase profile to minimize end-to-end image reconstruction error (Sec. 4). Our metasurface achieves an impressive phase control range exceeding 200°, spanning from 21° to -181°, while maintaining a penetration loss of less than -3dB for double penetration across W-band frequencies (77–81GHz). This innovative design provides unparalleled capabilities for high-resolution transmissive imaging in mmWave systems.
- We develop a signal-to-image diffusion model for image reconstruction. Theoretically, we prove that diffusion models are optimal estimators under mean squared error (MSE), outperforming traditional formulations such as LASSO in compressive sensing problems (Sec. 5). Empirically, our diffusion-based approach delivers exceptional imaging performance, reducing the root mean squared error (RMSE) of compressive sensing methods by half.
- We propose a robust and efficient channel estimation and recovery algorithm to address the challenges of mmWave channel variability over time and across diverse environments (Sec. 6). By mitigating channel dynamics, our method ensures consistent and reliable imaging performance, even in non-stationary conditions.
- We implement MIMSID and validate its performance through extensive experiments and testbed evaluations (Sec. 7 and 8). Our results demonstrate that the optimized metasurface reduces RMSE by up to 66% compared to configurations without a metasurface, while our diffusion model achieves a 62.5% RMSE reduction over ADMM-based reconstruction methods. Combined, MIMSID achieves a median RMSE of 0.061 on a 20cm × 20cm imaging plane with 1cm resolution.

2 Related Work

2.1 mmWave Imaging

mmWave imaging has attracted great attention for its high range resolution, penetration ability, and capability to function in optically occluded conditions, thus being studied extensively for autonomous driving [14, 17], object detection [23, 40], and gesture recognition [62]. In addition, mmWave is non-ionizing, making mmWave imaging favorable for human-involved applications, such as security checks [45] and body examinations [30]. Current mmWave imaging methods often apply Synthetic Array Radar (SAR) [14, 24, 30, 40, 45, 60] to extend the virtual aperture for high-resolution imaging but require mechanical scanning and large machinery sizes, making them hard to deploy in compacted spaces or unmanned inspection tasks. [24] applies a rotation SAR with a compacted size, but still requires machinery rotations, introducing extra risks of mechanical failures and power assumptions, hindering the potential of long-time operations. [43] achieves high-resolution imaging with handheld devices, but requires human control over handheld devices. Moreover, it applies Generative Adversarial Nets (GAN) to reconstruct high-resolution images, which are prone to model collapse. Several works apply metasurfaces to expand imaging aperture, but function in much lower operating frequencies, resulting in lower resolutions [15, 18].

Compared to the time- and space-consuming nature of SAR, Inverse Synthetic Aperture Radar (ISAR) leverages the object's movement and synthesize large aperture with the relative movements between the transceivers and the object, therefore is preferred in compact scenarios where the radar is static and the objects are moving uniformly, such as in tire-wear sensing [37] and assembly-line sensing [21, 36, 58]. However, ISAR requires knowing the target's speed or using other sensors to track its location, and is not suitable for imaging static objects or objects with arbitrary movement (*e.g.*, users passing through security checks).

Compared with existing work, MIMSID advances state-of-the-art by developing a diffusion-based imaging algorithm and jointly optimizing a passive mmWave metasurface and codebooks. By combining advanced neural networks with hardware optimization, MIMSID delivers a compact, high-resolution imaging system without requiring large antenna arrays or mechanical movement. This innovation not only improves imaging performance but also improves practicality, cost-efficiency, and robustness in the real world. environments.

2.2 mmWave Metasurface

Programmable metasurface. Programmable metasurfaces can change their transmissive or reflective properties in real-time by altering the voltage applied to unit cells. The unit cells are active elements respondent to voltage changes, such as positive intrinsic-negative (PIN) diodes [11, 19, 20], varactor diodes [7, 27, 52], liquid crystals [56], and RF switches [6, 51]. While programmable metasurfaces are more flexible and can be adjusted in real-time, the cost and fabrication difficulty of the active unit cells will increase dramatically as the design frequency increases and the corresponding wavelength shrinks. Thus, in our working frequency band of $77-81GHz$, we choose a PCB-printed passive metasurface to enable phase array expansion and single-shot mmWave imaging.

Passive metasurface. Passive metasurfaces are composed of passive unit cells that encode specific phase or amplitude-altering profiles [34, 38]. The unit cells often compose metallic patterns and dielectric layers, forming virtual dipoles responding to EM waves. Passive metasurfaces are cost- and power-efficient and easy to manufacture compared to active metasurfaces [4, 7, 11, 19, 20, 27], and thus are favored in many applications. Passive unit cells resonate at specific frequency bands depending on their physical structures and period sizes, enabling reflection [28], transmission [25, 33, 35, 50, 53], and refraction [32, 44, 57] functionalities. [28] proposes a reflective metasurface design with high reflectivity and near $2/pi$ phase control to improve mmWave network coverage. [32] presents refracting metasurfaces using phase gradient unit cells operating at $83GHz$. Reflective and refracting metasurfaces need to be mounted away from the transceivers to ensure the signals reflected from the surfaces reach the target, increasing the size and deployment. Moreover, reflective metasurfaces involve LOS and reflective paths, greatly increasing channel modeling complexity. Current studies on transmissive metasurfaces are mostly below the U band ($40G-60GHz$). [35] enhances Low Earth Orbit (LEO) satellite communication via stacked transmissive metasurfaces. However, there is great potential in using transmissive metasurfaces in the W-band frequencies ($70-110GHz$), encouraging us to seek a unit-cell design with a wide phase shifting range and low attenuation in the popular $77-81GHz$ automotive radar frequency band.

2.3 Diffusion Models for Image Generation

Traditional mmwave imaging methods with static radars are based on point detections with point-cloud measurements or computational photography methods that solve linear inverse problems. Point-cloud-based methods scan the objects through multi-directional measurements to reconstruct images from position-wise reflection signals. [62] uses a commodity WiFi antenna array with super-resolution algorithms to capture coarse human-body parts. [17] applies a CNN model to generate super-resolution images of vehicles from coarse point-cloud measurements. However, these Point-cloud-based methods are task-specific and low in resolution (above $10cm$ resolution). The linear inverse problem takes the single-shot measurement y of the object x and solves for $y = Ax + \epsilon$, where ϵ is the random noise. However, the LASSO-based ADMM solver is sub-optimal [2] while message-passing-based solvers (*e.g.*, AMP [10]) require strict assumptions on the sensing matrix and do not converge for general sensing matrices.

The recent development in diffusion models (DM) has drawn attention in various application fields for imaging generations [8, 22]. DMs progressively refine the image, employing an encoder-decoder network in each step, while traditional image generation models such as Generative Adversarial Networks (GANs) [16, 43] suffer from mode collapse and training instability, and often underperform DMs [9]. In more recent works, DMs are also used to generate RF signals for data expansion [5].

With the continuous development of DMs, numerous works have emerged interest in solving noisy linear inverse problems, *i.e.*, $y = Ax + \epsilon$ problems, with generative models [29]. [3] develops a Monte-Carlo-guided diffusion model to solve the ill-posed linear inverse problems. [22] uses DMs for solving linear inverse problems in the medical imaging domain, such as image-to-image translation, reconstruction, denoising, and anomaly detection. Nevertheless, we prove for the first time that conditional diffusion models are optimal estimators for linear inverse problems and develop a novel signal-to-image diffusion model that translates the reflected mmwave signals into corresponding object images.

3 mmWave Imaging Problem

Wireless imaging can be cast as solving the optimization problem $\min_{\boldsymbol{x}} |\boldsymbol{Ax} - \boldsymbol{r}|$, where we need to reconstruct an image $\boldsymbol{x}$ (flattened to

$N \times 1$) given a known $M \times N$ measurement matrix $\boldsymbol{A}$ (*e.g.*, wireless channel matrix) and a known $M \times 1$ received signal $\boldsymbol{r}$. M is the channel feature length and N is the total channel number. Each element in $\boldsymbol{x}$ represents the signal fraction reflected at that position.

mmWave uses Frequency Modulated Continuous Waves (FMCW) for imaging. Due to the limited sampling rate, only mixed signals (*i.e.*, TX reference signals × RX received signals after low-pass filters) instead of raw signals are available. According to [46], under slow modulation conditions, mixed signals are equivalent to raw signals. Therefore the above formulation applies to mixed signals.

The imaging reconstruction accuracy depends on the measurement matrix $\boldsymbol{A}$. Increasing $\boldsymbol{A}$'s rank enhances the accuracy of image reconstruction, which depends on the number of transmitter and receiver antennas and/or the effective bandwidth for measurements. However, increasing the number of antennas is not only costly but also limited by an existing upper bound on antenna numbers available in COTS mmWave modules. In addition, the effective bandwidth is also limited by the physical design of the mmWave module.

Next, we will introduce our metasurface optimized for mmWave imaging. Then we will explain the intuition behind replacing traditional compressive sensing with diffusion networks and introduce our signal-to-image diffusion model for imaging reconstruction.

4 Metasurface Design

4.1 Background on mmWave Metasurface

Huygens Metasurfaces (HMS) are composed of sub-wave-length structures acting as arrays of dipole components that manipulate the phase and amplitude of the transmitted or reflected waves, and therefore can be seen as antenna elements to form a massive phased array. HMS can significantly enhance the degree of freedom in both spatial and frequency domains and improve the complexity of signal channels. This physical enhancement in signal diversity cannot be achieved by algorithms alone. The system cannot reliably reconstruct missing details if the raw data is insufficient or lacks the necessary information. This insight encourages us to design a metasurface to increase imaging resolution and improve imaging performance. A metasurface design consists of (i) a microscopic design to determine the structure of unit cells, and (ii) a macroscopic design to determine the entire metasurface, including various parameters associated with each unit cell at each position to achieve the desired phase map.

4.2 Metasurface-Based mmWave Imaging Problem Formulation

Consider placing an mmWave metasurface between the mmWave module and the target. The mmWave module transmits signals toward a target and receives the reflected signals, as shown in Fig. 2. Let $\boldsymbol{H}_{t,m}(i,j)$ denote the mmWave channel from the i-th mmWave transmitter antenna to the j-th metasurface cell. Similarly, we have $\boldsymbol{H}_{m,r}(j,k)$ denote the channel from the j-th metasurface cell to the k-th mmWave receiver antenna. Meanwhile, we define $\boldsymbol{H}_{m,o}(j,k)$ as the channel from the j-th metasurface cell to the k-th grid in the imaging area, and $\boldsymbol{H}_{o,m}(k,j)$ as the channel from the k-th grid in the imaging area to the j-th metasurface cell. $\boldsymbol{w}$ denotes the transmitters' beamforming codewords for the design frequency.

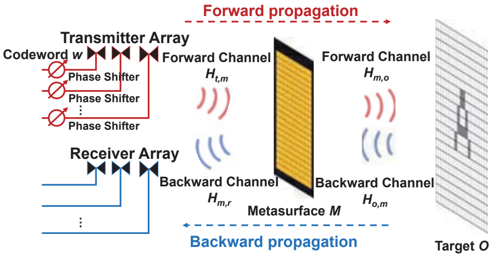


Figure 2: Channel modeling for mmWave imaging with metasurface.

The mmWave signal goes through the transmitter antennas' beamforming, the metasurface's modification, and arrives at the target, where it is reflected and goes through the metasurface in the reverse direction. It finally arrives at the mmWave receiver antennas and goes through their receiving processing. Let $\boldsymbol{r}$ denote the received signal at the receiver. We have:

$$\boldsymbol{r} = \boldsymbol{H}_{m,r}\boldsymbol{M} \cdot \boldsymbol{H}_{o,m}\boldsymbol{x} \cdot \boldsymbol{H}_{m,o}\boldsymbol{M} \cdot \boldsymbol{H}_{t,m}\boldsymbol{w} + \boldsymbol{n}, \tag{1}$$

where $\boldsymbol{M}$ is the modification to the signals introduced by the metasurface, $\boldsymbol{w}$ is the codeword (or the initial phase shift) of the transmitter, $\boldsymbol{n}$ is the random noise, and $\cdot$ denotes the dot product. A dot product captures the interactions between the metasurface $\boldsymbol{M}$ and the incoming signal because each metasurface cell changes the signals independently regardless of the signal's path. For the same reason, a dot product is used to model the impact of an object on the incoming signal.

Through simplification, Eq. 1 can be reduced to as follows:

$$\begin{aligned} \boldsymbol{r}_m &= \boldsymbol{H}_{m,r} diag(\boldsymbol{M}) \boldsymbol{H}_{o,m} diag(\boldsymbol{H}_{m,o}\boldsymbol{M} \cdot \boldsymbol{H}_{t,m}\boldsymbol{w})\boldsymbol{x} + \boldsymbol{n} \\ &= \boldsymbol{A}_m(\boldsymbol{M}, \boldsymbol{w})\boldsymbol{x} + \boldsymbol{n}, \end{aligned} \tag{2}$$

where $\boldsymbol{r}_m$ is the combined reflection signal at the receiver, $\boldsymbol{A}_m(\boldsymbol{M}, \boldsymbol{w})$ is the measurement matrix for the mixed signal channels. Note that $\boldsymbol{A}_m(\boldsymbol{M}, \boldsymbol{w})$ is solely determined by the relative positions between the metasurface, the transceiver, and the predefined virtual image plane, and thus is independent of real-world environments.

To infer an image of our target $\boldsymbol{x}$, we search for $\boldsymbol{x}$ that satisfies the above constraints. All the other variables in Eq. 2 are known: the passive metasurface $\boldsymbol{M}$ is fixed and known in advance, and the channel matrices, $\boldsymbol{H}_{t,m}$, $\boldsymbol{H}_{m,o}$, $\boldsymbol{H}_{o,m}$, $\boldsymbol{H}_{m,r}$, can be calibrated in advance or derived based on the relative position among the transceivers, metasurface $\boldsymbol{M}$, and the imaging area.

To support high-resolution imaging reconstruction, we need $\boldsymbol{A}$ to have a high rank. Without a metasurface, the rank of $\boldsymbol{A}$ is determined by the number of frequencies used for measurements and the number of transmitter and receiver antennas. Assuming we use a sufficient number of frequencies (or sufficient bandwidth as for FMCW chirps), the number of antennas puts the upper bound on $\boldsymbol{A}$'s rank. Since it is expensive to use a large number of antennas in terms of hardware cost, power, and computation, we develop a passive metasurface to expand to the phased array and enhance the performance of the mmWave imaging system.

4.3 Microscopic design

4.3.1 Challenges. The unit cells of the HMS contain metallic patterns with rings, lines, and gaps serving as inductance and capacitance. The metallic patterns and dielectric intermediates form dipole units that excite electromagnetic resonance and achieve efficient

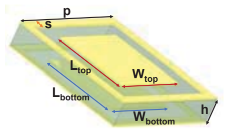


Figure 3: Unit cell geometry.

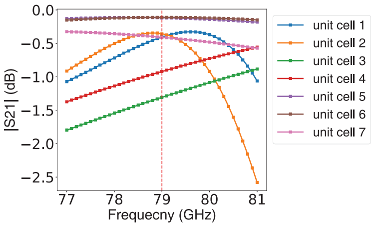


Figure 4: Transmission coefficient of unit cells.

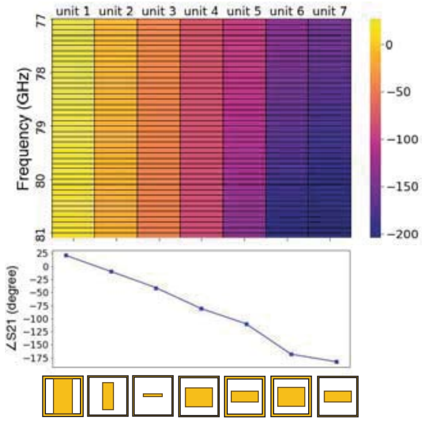


Figure 5: Phase manipulation via unit cell structure change.

transmission or reflection. Existing works on passive mmWave metasurfaces focus on reflective metasurfaces or narrow-band metasurfaces. While achieving high reflectivity is easier by adding a metal plate as the last layer [28], achieving high transmittance is more challenging. Without careful design, metasurfaces can significantly attenuate the signals going through them. Moreover, the large bandwidth of mmWave radars challenges us to achieve high transmittance across a wide range of frequency bands (*e.g.*, $4GHz$). While transmissive metasurfaces have been studied extensively below $40GHz$, few works exist on transmissive metasurfaces at W-band frequencies (75- 110GHz). The small wavelength in W-band mmWave makes it challenging to design and manufacture the metasurface.

HMS cells show strong phase change responses around their resonant frequencies and tend to reverse the transmittance responses (*i.e.*, a normally reflective metal pattern shows transmissive behavior, and vice versa). To achieve effective phase manipulation while maintaining high transmittance at a wide frequency range, the metasurface needs to show electromagnetic resonance across a broad frequency band. This challenges us to adopt low-Q resonance unit cells at W-band design frequencies.

4.3.2 Unit cell geometry. We redesign a rectangular pattern with low-frequency wide band resonance responses (low-Q resonance) [25] to achieve high transmittance and wide phase manipulation range at $77-81GHz$, as shown in Fig. 3. The unit cell is composed of a single dielectric layer sandwiched between two rectangular ambulatory plane metal layers. Each metallic pattern comprises an outer square ring and an inner rectangle patch. The period of meta-cells, denoted as p, is set to $1.5mm$, which is around half the wavelength of the design frequency $79GHz$. s is the width of the outer ring, and $h = 0.25mm$ is the thickness of the dielectric substrate. $W_{top,bottom}$ and $L_{top,bottom}$ are the widths and lengths of the rectangle patches, respectively.

The inductance and capacitance of the unit cell can be controlled by altering the width and lengths of the metallic structure, resulting in changes in phase and transmittance response. However, the subwavelength sizes of the unit cells significantly limit the design space to achieve an abundant variety of phase modulations. To achieve a larger design space in limited parameter selection ranges, we add more degrees of freedom by adjusting the width of the outer ring in addition to altering the shapes of the inner rectangles. At last, a wide range of phase modulation can be achieved by carefully selecting s, $W_{top,bottom}$, and $L_{top,bottom}$ pairs while maintaining high transmittance, forming a discrete collection of phase change responses for wavefront manipulation.

We use High-Frequency Structure Simulation (HFSS) to select effective geometry structures and determine the relationship between structural hyperparameters and phase modulation angles. We use the transmission parameter $S21$ as the frequency response of the meta-cell, where $\angle S21$ represents the phase modulation response and $\mid S21 \mid$ represents the transmittance of the meta-cell. We scan $W_{top,bottom}$ and $L_{top,bottom}$ from $0.1mm$ to $1.3mm$, and s from $0.05mm$ to $0.1mm$, while other hyperparameters are fixed. The transmittance response across $77-81GHz$ is shown in Fig. 4, where 7 selected unit cells are shown at random for clear representation. The phase response for 7 selected unit cells across the $77-81GHz$ frequency band and at the design frequency $79GHz$ is shown in Fig. 5. Our results show our single-layer meta-cells can support a phase modulation range of over 200 degrees with a transmission coefficient above $-2.5dB$ at the $77-81GHz$ frequency band and above $-1.5dB$ in our efficient working bandwidth(*i.e.*, 3800 MHz starting from 77GHz). Note that all unit cells present no significant peaks or valleys in transmittance across $77-81GHz$, indicating their low-Q resonance characteristics [25].

Although the phase response of each cell changes as the incident-wave frequency alters, the phase response change of each cell is approximately linear, and all cells share a similar slope in phase changes. As we only use the relevant phase changes (*i.e.*, $\Delta\angle S21$) to construct the phase map, we only care about the phase response span of all unit cells under the same incident-wave frequency. The transmission property is identical in both ways ($S21$ and $S12$), doubling the phase modulation ability and maintaining a penetration loss above $-3dB$. As shown in Fig. 5, the geometric parameter pairs and the transmissive phase responses are in one-to-one correspondence, indicating a mapping between geometry structures and phase response patterns. We select meta-cell parameters corresponding to phase response from $-180°$ to $20°$ in $5°$ interval. These meta-cells are arranged to form a phase map to manipulate transmissive wavefront, as described in Sec. 4.4.

4.4 Macroscopic Design

Our goal for the metasurface is to enhance imaging reconstruction. According to Sec. 4.2, while other parameters are determined by positions, we optimize the metasurface phase map $\boldsymbol{M}$ and transmitter's codewords $\boldsymbol{w}$ to realize this goal. We apply an end-to-end

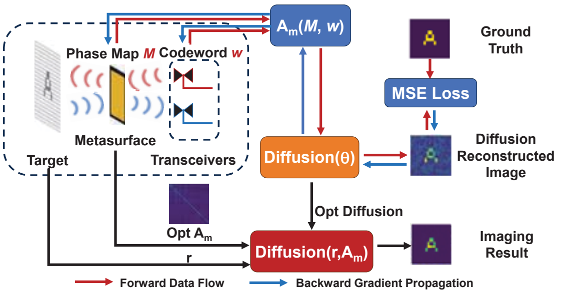


Figure 6: End-to-end phase map optimization

optimization framework to jointly optimize a signal-to-image diffusion network (described in Sec. 5) and the measurement matrix $\boldsymbol{A_m}$, as shown in Fig. 6.

The measurement matrix is determined by the metasurface phased map $\boldsymbol{M}$ at the design frequency of $79GHz$ and the phase offset codebook $\boldsymbol{w}$ for the transceivers. By iteratively optimizing imaging performance and the measurement matrix, we can obtain the optimized metasurface phase map $\boldsymbol{M_{opt}}$ and its corresponding codebook $\boldsymbol{w_{opt}}$, which greatly increases channel diversity and enhances imaging reconstruction performance.

We first initialize the measurement matrix $\boldsymbol{A_m}$ by randomly initializing $\boldsymbol{M}$ and $\boldsymbol{w}$. To ensure faster convergence and better $\boldsymbol{A_m}$ construction, we minimize the coherence of the matrix to increase the rank of $\boldsymbol{A_m}$.

$$\min_{\boldsymbol{M},\boldsymbol{w}} L = \sum_{i \neq j} C_{ij} \tag{3}$$

where $C_{ij} = \boldsymbol{A_m}(\boldsymbol{M}, \boldsymbol{w_i})^T \boldsymbol{A_m}(\boldsymbol{M}, \boldsymbol{w_j})$ computes the correlation between the i-th and j-th rows in the measurement matrix. As $\boldsymbol{A_m}$ is a function of $\boldsymbol{w}$ and $\boldsymbol{M}$, the final recovered image is a function of $\boldsymbol{w}$, $\boldsymbol{M}$, and θ, where $\theta = \{\theta_1, \theta_2, \theta_3\}$ is the trainable parameters of our *Sig2Img Diffusion* introduced in Sec. 5.3. We adopt the following end-to-end loss to optimize these parameters:

$$\ell_1(\boldsymbol{M}, \boldsymbol{w}, \theta) = \frac{1}{N|\mathcal{D}|} \sum_{\boldsymbol{x}_i \in \mathcal{D}} \|\boldsymbol{x}_i - \hat{\boldsymbol{x}}_i(\boldsymbol{M}, \boldsymbol{w}, \theta)\|_2^2,$$

where $\boldsymbol{x}_i$ is the ground-truth image from the training dataset $\mathcal{D}$ described in Sec. 5.2 consisting of 24900 images, $\hat{\boldsymbol{x}}_i$ is the estimated image, and N is the number of grids in the imaging plane. $\boldsymbol{M}$ is restricted into the phase modulation range described in Sec. 4.3.2.

We alternately optimize the metasurface parameters (*i.e.*, $\boldsymbol{A_m}$ and $\boldsymbol{w}$) and the diffusion model. The first step optimizes $\boldsymbol{M}$ and $\boldsymbol{w}$ and fixes θ while the second step optimizes θ and fixes $\boldsymbol{M}$ and $\boldsymbol{w}$. The two steps iterate until convergence. The optimization is done by backpropagation in Pytorch.

To meet the unit-cell design and physical constraints, the optimized phase map $\boldsymbol{M_{opt}}$ is mapped to the nearest feasible set from the selection described in Sec. 4.3.2, and $\boldsymbol{w_{opt}}$ is discretized into a 6-bit phase shifting profile to apply to the mmWave module.

5 Diffusion Models for Image Reconstruction

5.1 Background on Diffusion Models

Image generation has been a hot topic due to its diverse applications. Generative Adversarial Networks (GANs) [16, 43] gained popularity for their impressive capabilities in generating high-quality images. However, they are prone to model collapse (*i.e.*, generating less diverse samples, and the training is highly unstable).

Recently, Diffusion Models (DMs) have broken the long-standing dominance of GANs [9] as they elegantly addressed the above issues. DMs are inspired by the diffusion process in statistical physics and are characterized by forward and reverse processes. The forward process, occurring over a time interval from 0 to T, incrementally transforms an image into Gaussian noise by following the gradient of the image distribution. Let $\boldsymbol{x}_t$ represent the state of the data point at time t ($\boldsymbol{x}_0$ as the ground-truth image); the forward process systematically introduces noise to the data by following a predefined noise schedule given by $\boldsymbol{x}_t = \sqrt{\alpha_t}\boldsymbol{x}_0 + \sigma_t \boldsymbol{\epsilon}_t$, where $\alpha_t \in [0, 1]$ is monotonically decreasing with t, $\sigma_t = \sqrt{1 - \alpha_t}$. In contrast, the reverse process, from T back to 0, reconstructs the image from the noise. The reverse process is obtained by the equation:

$$\boldsymbol{x}_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}}\boldsymbol{x}_t + \sigma_t^2(e^h - 1)\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) \tag{4}$$

where p_t is the distribution of $\boldsymbol{x}_t$, $\lambda_t = \frac{1}{2}\log(\frac{\alpha_t}{1-\alpha_t})$, $h = \lambda_{t-1} - \lambda_t$.

In statistical physics, the expression of $\nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t)$ (*i.e.*, score function) is obtained in closed form while this is intractable for images. As a result, the key to DM is to use neural networks to approximate the score function, *i.e.*, to train an NN such that $\text{NN}(\boldsymbol{x}_t, t) \approx -\sigma_t \nabla_{\boldsymbol{x}_t} \log p_t(\boldsymbol{x}_t) = \boldsymbol{\epsilon}_t$ [48]. During inference, DMs start $\boldsymbol{x}_T$ with Gaussian noise, executing Eq (4) by replacing the score function with the neural network. The generated image is $\boldsymbol{x}_0$.

5.2 Image Reconstruction: From Compressive Sensing to Conditional Diffusion

In this subsection, we discuss the rationale behind using diffusion for image reconstruction. The key observation is that DMs are optimal in terms of MSE while traditional formulations like LASSO are not. We first follow [31] to introduce the compressive sensing methods from a probability perspective. The conditional probability of measurement is a Gaussian distribution, i.e., $p(\boldsymbol{r}|\boldsymbol{x}) \sim \exp(-\frac{1}{2\sigma^2}\|\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}\|_2^2)$. The joint probability is $p(\boldsymbol{x}, \boldsymbol{r}) = p(\boldsymbol{r}|\boldsymbol{x})p_0(\boldsymbol{x})$, where $p_0(\boldsymbol{x})$ is the distribution of images. Then we consider a minimum mean square error estimator (MMSE) to recover $\boldsymbol{x}$ from $\boldsymbol{r}$:

$$\begin{aligned}
\boldsymbol{x}^* &= \operatorname*{argmin}_{\hat{\boldsymbol{x}}} \mathbb{E}[\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2], \boldsymbol{x} \sim p(\boldsymbol{x}|\boldsymbol{r}) = \mathbb{E}[\boldsymbol{x}] \\
&= \int \boldsymbol{x} p(\boldsymbol{x}|\boldsymbol{r}) \mathrm{d}\boldsymbol{x} = \int \boldsymbol{x} \frac{p(\boldsymbol{r}|\boldsymbol{x})p_0(\boldsymbol{x})}{p(\boldsymbol{r})} \mathrm{d}\boldsymbol{x} = \frac{1}{p(\boldsymbol{r})} \int \boldsymbol{x} p(\boldsymbol{r}|\boldsymbol{x}) p_0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \\
&= \frac{1}{p(\boldsymbol{r})} \int \boldsymbol{x} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}\|_2^2\right) p_0(\boldsymbol{x}) \mathrm{d}\boldsymbol{x},
\end{aligned} \tag{5}$$

By taking $p_0(\boldsymbol{x}) = \exp(-\lambda\|\boldsymbol{x}\|_1)$ and replacing the integral with the maximum operation, one may recover the classic formulation of LASSO:

$$\begin{aligned}
\boldsymbol{x}^* &= \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{r}) = \operatorname*{argmax}_{\boldsymbol{x}} \frac{p(\boldsymbol{x}|\boldsymbol{r})p_0(\boldsymbol{x})}{p(\boldsymbol{r})} = \operatorname*{argmax}_{\boldsymbol{x}} p(\boldsymbol{x}|\boldsymbol{r})p_0(\boldsymbol{x}) \\
&= \operatorname*{argmax}_{\boldsymbol{x}} \exp\left(-\frac{1}{2\sigma^2}\|\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}\|_2^2\right) \exp(-\lambda\|\boldsymbol{x}\|_1) \\
&= \operatorname*{argmin}_{\boldsymbol{x}} \frac{1}{2}\|\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \lambda\sigma^2\|\boldsymbol{x}\|_1.
\end{aligned} \tag{6}$$

Due to the replacement of mean by max, LASSO is not optimal in MSE even if $\exp(-\lambda\|\boldsymbol{x}\|_1)$ is the ground-truth data distribution (Fig. 2 of [31]). To solve Eq. 5, the most effective approaches are

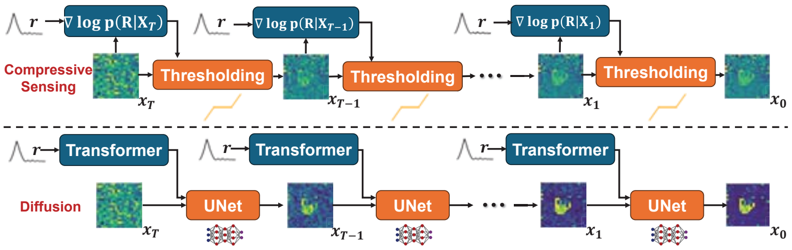


Figure 7: Comparison of the proposed diffusion methods and conventional compressive sensing methods.

approximate message passing [1]:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{x}_t - \nabla_{\boldsymbol{x}_t}\frac{1}{2}\|\boldsymbol{r}-\boldsymbol{A}\boldsymbol{x}_t\|_2^2 + \boldsymbol{c}_t = \boldsymbol{x}_t + \boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{r}) + \boldsymbol{c}_t \\ \boldsymbol{x}_{t-1} &= \eta_t(\boldsymbol{y}_t), \end{aligned} \tag{7}$$

where $\boldsymbol{c}_t$ is Onsager correction term [10] and the thresholding function is given by

$$\eta_t(\boldsymbol{y}_t) = \int \boldsymbol{x}\exp\left(-\frac{1}{2\sigma_t^2}\|\boldsymbol{x}-\boldsymbol{y}_t\|_2^2\right)p_0(\boldsymbol{x})\mathrm{d}\boldsymbol{x}, \tag{8}$$

For sparse problems (i.e., $p_0(\boldsymbol{x}) = \exp(-\lambda\|\boldsymbol{x}\|_1)$, it is the soft-thresholding operator [10], i.e., $\eta_t(\boldsymbol{y}_t) = \mathrm{sign}(\boldsymbol{y}_t)\cdot\max(|\boldsymbol{y}_t| - \lambda_t\sigma_t)$. One important connection between the optimal message passing methods Eq. 7 and diffusion models are given in the following proposition and the derivation follows Eq. 11 - Eq. 16 of [61] and compared in Fig. 7.

PROPOSITION 5.1. *Given* $\boldsymbol{y}_t = \sqrt{\alpha_t}\boldsymbol{x}^* + \sigma_t\epsilon_t$, *then we have* $\eta_t(\boldsymbol{y}_t) = \frac{1}{\sqrt{\alpha_t}}(\boldsymbol{y}_t + \sigma_t^2\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t))$, *where* $\sqrt{\alpha_t}, \sigma_t$ *and* $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$ *are defined in Eq. 4.*

As discussed in the previous subsection, the key to diffusion is to use a neural network to approximate $\nabla_{\boldsymbol{x}_t}\log p_t(\boldsymbol{x}_t)$. Therefore, in this paper, we use diffusion models to approximate $\eta_t(\boldsymbol{y}_t)$. We will not lose the optimality as the additional identity mapping and scaling by $1/\sqrt{\alpha_t}$ is easy to learn. Specifically, we initialize the image as Gaussian noise and use the following updates mimicking Eq. 7:

$$\begin{aligned} \boldsymbol{y}_t &= \mathrm{Signal\_NN}(\boldsymbol{A}, \boldsymbol{r}) \\ \boldsymbol{x}_{t-1} &= \mathrm{Diff\_NN}(\boldsymbol{x}_t, \boldsymbol{y}_t) \end{aligned} \tag{9}$$

where Signal_NN is a neural network for handling the received signal $\boldsymbol{r}$ under channel matrix $\boldsymbol{A}$ which encodes the signal inputs $\boldsymbol{A}$ and $\boldsymbol{r}$ into representation embeddings and Diff_NN is a diffusion neural network conditioned by the embeddings to execute Eq. 4. In the next subsection, we will discuss the neural architecture of Signal_NN and Diff_NN, as well as the training method in detail. We name our algorithm Eq. 9 as *Sig2Img Diffusion*. Our *Sig2Img Diffusion* aims at learning the intrinsic relationship between the received signal $\boldsymbol{r}$ and channel matrix $\boldsymbol{A}$ and performs image reconstruction under the guidance of received signals.

5.3 Implementation of Sig2Img Diffusion

Training data generation. The training of diffusion models typically requires a large amount of paired data consisting of the images and received signals, which is not always available in mmWave sensing. Instead, we propose to synthesize image-signal data pairs using the channel matrix $\boldsymbol{A} \in \mathbb{R}^{N_{samples}\times 400}$, where $N_{samples}$ is the number of signal sample points in each channel. Specifically, we generate 4900 ground-truth images according to our use cases (*e.g.*, weapon detection, pipeline speculation, digits, and alphabets) together with 20,000 images in Fashion-MINST [59] to construct the image dataset $\mathcal{D}_{\mathrm{image}}$. For each image $\boldsymbol{x}_i \in \mathcal{D}_{\mathrm{image}}$, we obtain the corresponding signal $\boldsymbol{r}_i = \boldsymbol{A}_i\boldsymbol{x}_i$, where $\boldsymbol{A}_i \in \{\boldsymbol{A}_k\}_{k=1}^{36}$ is drawn from a simulated $\boldsymbol{A}$ set with various channel setups. As our *Sig2Img Diffusion* takes the received signal $\boldsymbol{r}$ and the corresponding channel matrix $\boldsymbol{A}$ pairs as inputs, the synthetic dataset consists of each signal-image pair $\mathcal{D} = \{\boldsymbol{x}_i, (\boldsymbol{r}_i, \boldsymbol{A}_i)\}_{i=1}^{24900}$, where the received signal $\boldsymbol{r}_i \in \mathbb{R}^{N_{samples}}$, the corresponding channel matrix $\boldsymbol{A}_i \in \mathbb{R}^{N_{samples}\times 400}$ and the image is 20×20, *i.e.*, $\boldsymbol{x}_i \in \{0,1\}^{400}$.

The neural architecture of our signal-to-image DM is built on a seminal text-to-image DM named Stable Diffusion [42]. In stable diffusion, the text is first encoded as a feature $\boldsymbol{y}_t$. Then a UNet is employed to take $\boldsymbol{y}_t$ and noisy image $\boldsymbol{x}_t$ as input and output $\boldsymbol{x}_{t-1}$.

Signal_NN. Signal_NN extracts the features of received signals $\boldsymbol{r}$ and their corresponding channel matrices $\boldsymbol{A}$, aiming to learn the general relationship between $\boldsymbol{r}$ and $\boldsymbol{A}$. Intuitively, if the signal features are more similar to the corresponding image features, it will be easier for the diffusion models to recover the image. Therefore, we first align the representation of the image and the corresponding signal. We compress the representation of $\boldsymbol{A}$ using a 2-layer 1D CNN and concatenate it with $\boldsymbol{r}$ to construct a paired signal vector $\boldsymbol{r_P}$. We employ a 1-layer 1D Transformer and a 2-layer 2D Transformer to respectively obtain the features of paired signals and images:

$$\boldsymbol{y}_{i,S} = \text{1D-Transformer}_{\theta_1}(\boldsymbol{r_P}_i), \quad \boldsymbol{y}_{i,I} = \text{2D-Transformer}_{\theta_2}(\boldsymbol{x}_i), \tag{10}$$

where $\boldsymbol{y}_{i,S}, \boldsymbol{y}_{i,I} \in \mathbb{R}^{400}$, $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ denote the weights of neural networks. We adopt the CLIP loss (*i.e.*, Fig. 3 in [39]) as the loss function. It encourages $\boldsymbol{y}_{i,S} = \boldsymbol{y}_{j,I}$ if $i = j$ and decreases the similarity between $\boldsymbol{y}_{i,S}$ and $\boldsymbol{y}_{j,I}$ for $i \neq j$. The weights of 1D and 2D-Transformer are trained by backpropagating the loss function to $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ with AdamW optimizer [26].

Diff_NN. Diff_NN takes the features of the signal pair $\boldsymbol{r_P}$ and the noisy image $\boldsymbol{x}_t$ as the input and outputs the denoised image $\boldsymbol{x}_{t-1}$. For each image in training dataset $(\boldsymbol{x}_i, \boldsymbol{r_P}_i) \in \mathcal{D}$, it first goes through the forward diffusion process

$$\boldsymbol{x}_{i,t} = \sqrt{\alpha_t}\boldsymbol{x}_i + \sigma_t\epsilon_t. \tag{11}$$

As the input and output dimensions are identical to stable diffusion, we adopt the diffusion network in stable diffusion (*i.e.*, UNet) as the backbone of Diff_NN:

$$\hat{\epsilon}_{i,t}(\boldsymbol{r_P}_i, \boldsymbol{x}_{i,t}) = \mathrm{UNet}_{\theta_3}(\text{1D-Transformer}_{\theta_1}(\boldsymbol{r_P}_i), \boldsymbol{x}_{i,t}). \tag{12}$$

where $\boldsymbol{\theta}_3$ is the UNet's weights. The training loss is to minimize the prediction error given by the UNet:

$$\sum_{i\in|\mathcal{D}|}\sum_t \|\hat{\epsilon}_{i,t}(\boldsymbol{r_P}_i, \boldsymbol{x}_{i,t}) - \epsilon_t\|_2^2. \tag{13}$$

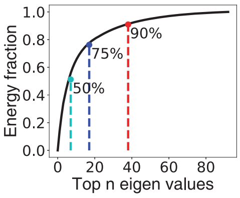
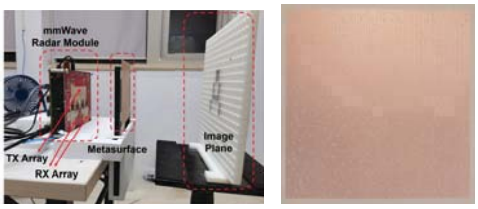


Figure 8: Eigen value distribution of ΔA



(a) System prototype

(b) Frontview of the fabricated metasurface

Figure 9: Photograph of the fabricated prototype.

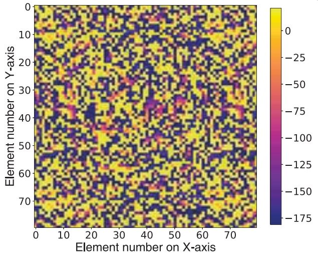


Figure 10: Metasurface optimized phase profile

The weights of UNet are learned by backpropagating the loss function to $\boldsymbol{\theta}_3$ with the AdamW optimizer. The training is finished on one RTX 3060 desktop in 9 hours. During the inference, given a new paired signal vector $\boldsymbol{r_P}$, we initialize $\boldsymbol{x}_T$ with Gaussian noise. The image is recovered by the inverse diffusion process, given by

$$\boldsymbol{x}_{t-1} = \sqrt{\frac{\alpha_{t-1}}{\alpha_t}}\boldsymbol{x}_t - \sigma_t(e^h - 1)\hat{\epsilon}_t(\boldsymbol{r_P}, \boldsymbol{x}_t). \tag{14}$$

The recovered image from the signal is $\boldsymbol{x}_0$.

6 Efficient Channel Updating

MIMSID requires paired $\boldsymbol{A}$ and $\boldsymbol{r}$ as the inputs of *Sig2Img Diffusion*. However, the channels in $\boldsymbol{A}$ may change over time and in different environments. Therefore, $\boldsymbol{A}$ requires periodical updating. A natural way to update the channels is to measure the reflected channel at each point in the imaging plane (*i.e.*, recalibration). To reduce the measurement overhead, we develop a compressive sensing method using a few template images to recover updated $\boldsymbol{A}$.

Let $\boldsymbol{X} \in \mathbb{R}^{144\times k}$ denote k ($k = 30$ in experiments) flattened template images of 12×12, which are known in advance. We place each template image at the target plane and measure the received mmWave signal. Our goal is to infer the measurement matrix $\boldsymbol{A}$ (or its change $\Delta\boldsymbol{A}$) based on the known template images $\boldsymbol{x}$ and received signals $\boldsymbol{r}$, where $\boldsymbol{r} = \boldsymbol{A}\boldsymbol{x}$. This problem is massively underconstrained when the number of template images is small.

We need to leverage the special structure in the measurement matrix. Compressive sensing typically assumes the unknowns are sparse or low-ranked. Carefully examining the distribution of eigenvalues in the channel matrix shows that $\boldsymbol{A}$ and $\Delta\boldsymbol{A}$ are neither low rank nor sparse. For example, as shown in Fig. 8, it takes 33 eigenvalues (23% eigenvalues) to account for 90% of the energy of $\Delta\boldsymbol{A}$ with a 144 measurement matrix (*i.e.*, 12×12 image plane).

Inspired by [49], we use the spectral profile as the regularization term. Instead of enforcing a strict low-rank condition, we use a given distribution of spectral profile as the input. This is a more flexible constraint and can accommodate different types of distributions. Specifically, we formulate the problem as follows:

$$\min_{\boldsymbol{A}} \frac{1}{2}\|\boldsymbol{r} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + I(\boldsymbol{A}, P). \tag{15}$$

Here, $I(\boldsymbol{A}, P)$ denotes the indicator function, and $I(\boldsymbol{A}, P) = 0$ if the new measurement $\boldsymbol{A}$ satisfies the spectral profile P. Otherwise $I(\boldsymbol{A}, P) = \infty$. We use the standard ADMM algorithm [49] to obtain the solutions (with the index from T to 0). The updates are given by

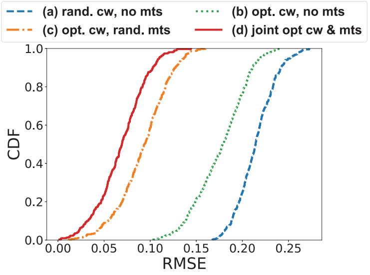


Figure 11: Overall performance CDF.

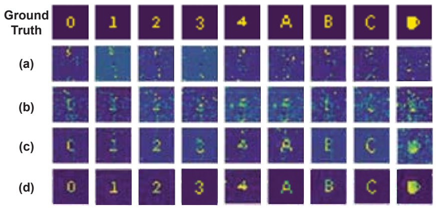


Figure 12: Overall performance examples.

$$\begin{aligned}
\boldsymbol{A}_{t-1} &= \left(\boldsymbol{Z}_t - \boldsymbol{N}_t/\mu + \boldsymbol{r}\boldsymbol{x}^T\right)(\boldsymbol{x}\boldsymbol{x}^T + \boldsymbol{I})^{-1},\\
\boldsymbol{Z}_{t-1} &= \text{Rescaling SVD}\left(\frac{1}{\mu}\boldsymbol{N}_t + \boldsymbol{A}_{t-1}\right),\\
\boldsymbol{N}_{t-1} &= \boldsymbol{N}_t + \mu(\boldsymbol{A}_{t-1} - \boldsymbol{Z}_{t-1}),
\end{aligned} \tag{16}$$

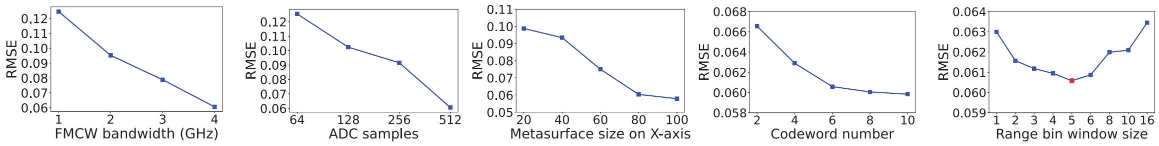
where $\boldsymbol{Z}_t, \boldsymbol{N}_t$ are auxiliary variables, and Rescaling SVD is Algorithm 1 in [49]. The above algorithm is general: we can apply it to estimate $\boldsymbol{A}$ or $\Delta\boldsymbol{A}$ (*i.e.*, the change in $\boldsymbol{A}$). Estimating $\Delta\boldsymbol{A}$ can be more efficient than estimating $\boldsymbol{A}$ if the change is small.

7 Prototype Implementation

7.1 Experimental Setup

Our system prototype is shown in Fig. 9. Our experimental setup consists of an mmWave radar module as FMCW transceivers, a transmissive metasurface, and a 3D-printed grid as the image plane. The metasurface is placed $10cm$ away from the transceiver board, with the centers of the metasurface and the transceiver array aligned. The passive static metasurface is easy to deploy in front of the radar, minimizing the risk of misalignment between the metasurface and the transceiver.

We use a commodity mmWave radar evaluation module MMWCAS-RF-EVM (Texas Instruments), which contains four cascaded AWR2243 FMCW transceiver chips and has 12 TXs and 16 RXs, forming a 192-channel virtual transceiver array. The TXs transmit FMCW chirps



(a) Bandwidths. (b) ADC samples. (c) Metasurface sizes. (d) Codeword numbers. (e) Range bin window sizes.

Figure 13: Effect of different design components on the imaging system.

with a slope of $95MHz/\mu s$ and a bandwidth of $3800MHz$, starting from $77GHz$. The RXs take 512 samples with a sampling rate of $20MHz$ in each chirp. Each TX transmits 6 unique chirps determined by the codewords in a Multi-Input-Multi-Output (MIMO) manner, forming a unique chirp profile of 6 chirps per transmit loop, which takes $0.24ms$ to transmit.

For imaging, we assume the target lies at a pre-defined image plane consisting of 20×20 grids, where each grid occupies $1cm\times1cm$. For ease of constructing various target shapes, we use small stainless steel cubes with an edge length $1cm$ as image pixels. Unless stated otherwise, the image plane is placed $30cm$ away from the transceiver board and uses $1cm$ resolution.

We perform a calibration process to obtain the mixed-signal measurement matrix following the modeling in Section 3. The calibration process constructs the measurement matrix by measuring the reflective signal responses of each image plane pixel. This calibration process can be automated using a 2D electric-controlled guide rail or a robot arm. We disable the boot-time calibrations of the mmWave radar to avoid random amplitude and phase jumping between module reboots [54].

7.2 Metasurface Fabrication

The metasurface is fabricated using standard photolithographic techniques on Rogers RO3003 laminates. RO3003 laminates offer excellent stability of low dielectric constant (Dk) under $77GHz$ mmWave radar frequencies, with a dielectric constant of $Dk = 3\pm0.04$ and loss tangent of $Df = 0.001$.The metasurface is composed of 80×80 unit cells and occupies $13cm\times13cm$. The phase profile of the metasurface is obtained via end-to-end optimization described in Section 4.4 and shown in Fig. 10. The cost of our metasurface is approximately \$275. With mass production, the cost could be reduced significantly (*i.e.*, \$0.04 each cell).

7.3 Signal Processing and Model Training

Real-signal processing: The received signals from the targets require signal processing to obtain representative features. We first convert the received signals to the frequency domain to extract range-related information. We perform range gating (*e.g.*, the distance 0-1m) to exclude the reflection from far objects. Our system periodically (*e.g.*, every hour) collects a background signal when there is no object to subtract the background interference. Then we detect the peak energy within the range of interest to locate the target. Since the imperfect hardware and near-field reflection will disperse the reflection to its adjacent range bins, we use multiple range bins near the imaging plane for imaging.

Diffusion model training: We train the signal-to-image diffusion model using synthesized training data as described in Section 5.3. The full training process takes 9 hours on one RTX 3060 desktop. Note that full training is done only once, and fine-tuning with fewer updated training samples can be done in 30 minutes.

Model inference: For model inference, we use the Denoising Diffusion Implicit Model (DDIM) [47] for skip sampling and fast inference. We tested the computational cost of model inference on an RTX 3060 desktop. The average inference time of our model is approximately 0.2 s per image for 20×20 pixel images.

8 Evaluation

8.1 Evaluation Methodology

Performance metric. The imaging performance of our system is measured via Root Mean Square Error (RMSE), which is defined as $RMSE = \sqrt{(est - gt)^2}$, where est is the estimated image and gt is the corresponding ground truth. Empirically, we consider a high-quality reconstruction performance to have RMSE<0.1 [12].

Baseline schemes. We evaluate the effectiveness of our imaging system by comparing 4 baseline setup schemes, as shown in Fig. 11, which are stated as follows. (a) Imaging system with random beamforming codewords and no metasurface. (b) Imaging system with optimized codewords and no metasurface. (c) Imaging system with optimized codewords and a metasurface with random phase profile. (d) Imaging system with the jointly optimized codewords and metasurface.

8.2 Overall Performance

Fig. 11 shows the Cumulative Distribution Function (CDF) of the RMSE of 4 schemes: (a) - (d). Fig. 12 shows some example images estimated using the 4 schemes. It is evident that the imaging system with jointly optimized codewords and metasurface (setup (d)) performs best, achieving a median RMSE of 0.0608. In contrast, without optimized codewords or the metasurface (setup (a)), one cannot generate any reasonable image. Scheme (d) outperforms scheme (a)-(c) in reducing RMSE by 72%, 66%, and 36.5%, respectively. Note that even a random metasurface (setup (c)) can greatly improve our system's performance, suggesting the benefit of metasurface in increasing the virtual phase array size and diversifying the beam patterns.

8.3 Microbenchmarks

In this section, we evaluate the impact of various design parameters under the optimal setup scheme (scheme (d)). Unless otherwise stated, the default system parameters are referred to those described in Sec. 7.1.

Vary valid FMCW bandwidth. The valid FMCW bandwidth is the fraction of bandwidth that can be sampled by the mmWave module in a mmWave chirp, defined as $B_{valid} = \frac{B_{total}Sample_{ADC}}{t_{ramp}r_{ADC}}$, where B_{total} is the total bandwidth of the FMCW chirp, $Sample_{ADC}$ is the number of samples captured by ADC, t_{ramp} is the chirp duration, and r_{ADC}

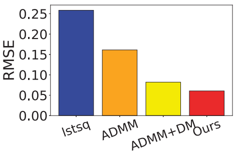


Figure 14: RMSE w.r.t. different imaging algorithms

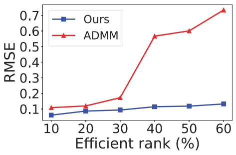
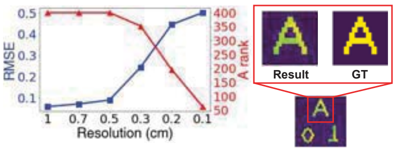


Figure 15: Imaging performance with different image sparsity



(a) (b)

Figure 16: Imaging under different resolutions. (a) RMSE vs. resolution. (b) 5mm resolution example.

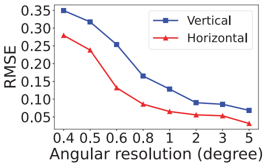


Figure 17: RMSE w.r.t. angular resolutions.

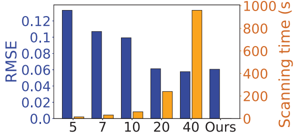


Figure 18: Comparison with SAR imaging.

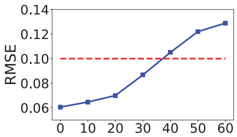


Figure 19: Impact of AoA

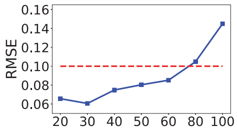


Figure 20: Impact of imaging distance

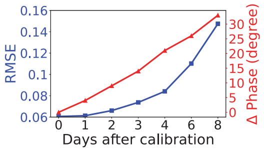


Figure 21: Consistency over time.

is the ADC sampling rate. The valid bandwidth determines the range resolution of the mmWave radar and hence affects the performance of our imaging system. In this work, we fix $t_{ramp} = 40\mu s$ and $r_{ADC} = 20MHz$ as described in Sec. 7.1, and evaluate the impact of various total chirp bandwidths and ADC samples.

Vary total bandwidth. Fig. 13(a) shows the impact of different total chirp bandwidths on the imaging performance. The median RMSE gradually decreases as the chirp bandwidth increases, from 0.125 with $1GHz$ bandwidth to 0.06 with $4GHz$ bandwidth. This observation encourages us to exploit the full $4GHz$ bandwidth provided by the radar module to achieve good performance.

Vary ADC samples. Fig. 13(b) shows the impact of different ADC samples. The median MSE declines from 0.125 with 64 samples per chirp to 0.06 with 512 samples per chirp. Similar to total chirp bandwidth, more ADC samples per chirp result in higher valid bandwidth, hence achieving better performance. Increasing the total chirp bandwidth and ADC sampling rate improves the effective bandwidth, which in turn improves imaging resolution.

Vary metasurface size. We evaluate the impact of different metasurface sizes, as shown in Fig. 13(c). The median RMSE decreases significantly from 0.098 with 20×20 metasurface to 0.06 with 80×80 metasurface. However, further increasing metasurface size to 100×100 unit cells yields a marginal improvement. Note that even a small metasurface of 20×20 cells increases the channel diversity and improves the imaging performance, compared with no metasurface (scheme (b)).

Vary codeword number Next, we evaluate the effect of different numbers of the phase offset codewords $\boldsymbol{w}$, as shown in Fig. 13(d). Similar to the impact of metasurface sizes, the median RMSE drops from 0.067 to 0.061 as the codeword number increases from 2 to 6. After the codeword number exceeds 6, the performance benefits are negligible.

Vary targeting window size. As the imaging plane is placed $30cm$ away from the mmWave transceiver, the peak energy lies in the range of interest of the target image plane. However, information from a single range bin is insufficient for effective imaging due to signal leakage across range bins. Fig. 13(e) plots the impact of applying different range bin window sizes around the target range bin. The best performance is achieved using a window size of 5 range bins. Too few range bins around the target range bin could lead to missing useful information due to signal leakage. Too large range bin windows may contain noise introduced by multipath and environmental objects, which also degrades the imaging performance and stability.

Performance of different imaging reconstruction algorithms. We further compare our imaging reconstruction method (*Sig2Img Diffusion*) with three other imaging reconstruction strategies on our testbed: (i) least square linear regression (lstsq), (ii) traditional ADMM, and (iii) denoising diffusion using ADMM outputs (ADMM + Diffusion). Fig. 14 shows that our signal-to-image diffusion method achieves the best performance due to the effectiveness of our diffusion approach. Compared with the traditional compressive sensing method ADMM, our approach automatically learns the image distribution and does not rely on the low-rank assumption, outperforming ADMM by 60%. Compared with the image-to-image diffusion approach, our signal-to-image diffusion learns the inherent relationship between the measurement matrix $\boldsymbol{A}$ and target signal $\boldsymbol{r}$ and exploits the direct relationship between reflected signals and targeted images, thereby achieving better image reconstruction by 25%.

8.4 Performance with Non-sparse Images

Traditional compressive sensing methods, such as ADMM, assume the target images to be sparse or low-ranked. However, the sparse constraint does not necessarily hold in real-world images. We compare the reconstruction performance of the ADMM algorithm and our signal-to-image diffusion network on images with different effective ranks, as shown in Fig. 15. Since the traditional compressive sensing method assumes low-rank conditions, it fails to reconstruct reasonable images after the rank exceeds a certain threshold. On the other hand, our signal-to-image diffusion automatically adopts the image distributions and is robust under different effective ranks.

8.5 Imaging with Higher Resolutions

Owing to the short wavelength of 77GHz mmwave, MIMSID is capable of performing accurate image reconstructions in sub-centimeter resolution. Fig. 16 (a) shows the performance of MIMSID under various resolutions below $1cm$. The optimal rank of $\boldsymbol{A}$ decreases as the resolution becomes more fine-grained. When the resolution is below the mmWave wavelength ($3.8mm$), RMSE increases dramatically. The maximum resolution for MIMSID to achieve high-quality imaging (*i.e.*, RMSE<0.1) is $5mm$, indicating MIMSID's ability to perform high-accuracy imaging under sub-centimeter resolution. Fig. 16 (b) shows an example at the $5mm$ resolution.

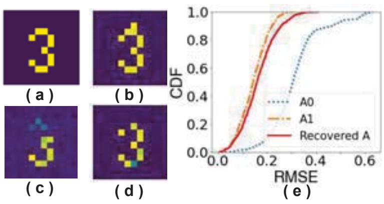


Figure 22: Recovered A performance.

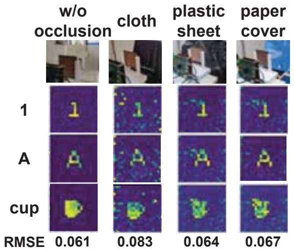


Figure 23: Impact of occlusion.

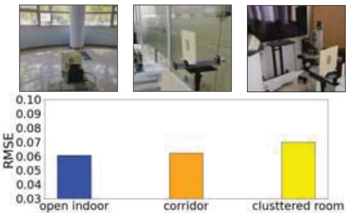
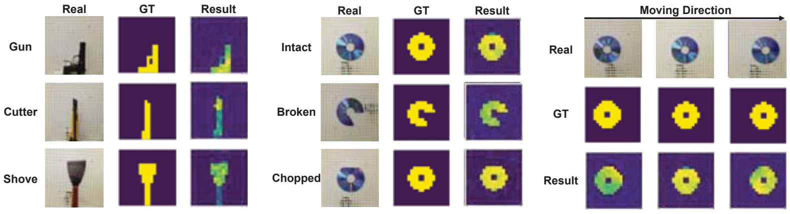


Figure 24: Impact of environment.



(a) Imaging for security checks. (b) Imaging for product inspections. (c) Simulation of moving object imaging.

Figure 25: Real-object Imaging.

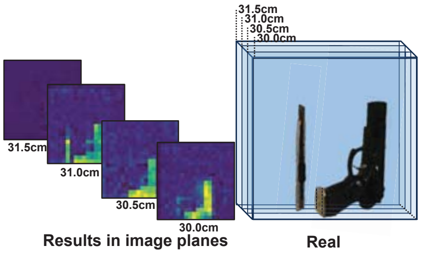


Figure 26: 3D imaging with multiple image planes.

8.6 Angular Resolution Analysis

To further evaluate the gain in the resolution of MIMSID, we experiment to show the performance of our system under different angular resolutions. We place two adjacent pixels on the image plane, either horizontally or vertically, and gradually vary their separation, hence adjusting the included angle between these pixels and the transceiver from 0.4° to 5°. Low RMSE indicates that our system is able to resolve the two pixels as distinct pixels. As shown in Fig. 17, the RMSE remains under 0.1 when the angular resolutions are larger than 0.8° horizontally and 2° vertically. The difference in the horizontal and vertical directions results in different antenna array sizes in the two directions. Compared with the MWCASRF-EVM radar without a metasurface (with 1° azimuth resolution and 15° elevation resolution), the gains in angular resolution are 125% horizontally and 750% vertically. The gain difference favoring the vertical direction is due to the end-to-end optimization framework for image reconstruction, which simultaneously enhances the horizontal and vertical resolutions of the system towards the same resolution.

8.7 Comparison with SAR

We compare MIMSID with the state-of-the-art mmWave imaging method, namely Synthetic Array Radar (SAR), as shown in Fig. 18. Our static imaging system, containing 86 virtual transceiver antennas in the azimuth direction and under a size of $15cm \times 15cm \times 15cm$, is comparable to a SAR imaging system with $20cm \times 20cm$ aperture size and $8 \times 192 \times 24$ virtual transceiver antennas, where 192 and 24 are the scanning steps in x and y direction, respectively. However, SAR systems require large deployment spaces and mechanical movements, while our prototype system is small and stationary, and can fit into compact assembly lines and narrow corridors. Moreover, as shown in Fig. 18, if the SAR system scans at $2cm/s$ [43, 60], the scanning time of a $20cm \times 20cm$ SAR is $240s$. If the object is moving in an assembly line, the long scanning time can result in image distortion or missing the object. In contrast, MIMSID scans a frame in less than $0.3ms$ and treats the slow-moving object (*e.g.*, $1m/s$) as a static object, thereby achieving good imaging quality.

8.8 Impact of Environment

MIMSID takes $\boldsymbol{A}$ and $\boldsymbol{r}$ as the input, eliminating the model's dependency on specific environments. However, owing to the propagation loss with reflection angles, distances, and occlusions, the SNR of $\boldsymbol{A}$ and $\boldsymbol{r}$ drops, resulting in worse imaging performance. In this section, we evaluate the impact of various environmental changes on our imaging system.

Impact of AoA. To evaluate the impact of different angles of arrival (AoAs), we place the image plane at different angles from the center of the mmWave module and perform recalibration each time. We vary the angle from 0° to 60°. Fig. 19 plots the median RMSE across a wide range of AoAs. The system's performance is robust across an AoA range of -40° to 40°. This shows our system can reconstruct images accurately even at low SNR and with angle misalignments with the predefined imaging plane.

Impact of imaging distance. Compared to acoustic imaging approaches [12], mmWave signals can achieve high-accuracy imaging at longer distances owing to their short wavelength and relatively low attenuation. We evaluate the imaging performance of different imaging distances. We place the imaging plane at varying distances from the center of the mmWave module and perform recalibration each time. Fig. 20 shows that our imaging system can accurately reconstruct high-quality images (RMSE < 0.1) up to $80cm$ with $1cm$ imaging resolution, well-suited for applications such as security checks in a corridor or product inspections beside an assembly line.

Impact of channel drifting. Due to the imperfection of the mmWave hardware, signal channels (phase and amplitude) may drift over time. However, as shown in Fig. 21, our imaging system can perform high-accuracy image reconstruction (RMSE<0.1) over 5 days. This indicates our imaging system is consistent over a long time before recalibration to update $\boldsymbol{A}$. Moreover, we can apply efficient channel estimation (Sec. 6) to update the measurement matrix $A_{recover}$ using a few samples (*e.g.*, 30 samples for a 144 image plane). Our estimated channel $A_{recover}$'s performance (Fig. 22 (d)) is close to the ground truth channel A_1 (Fig. 22 (b)), much better than the channel measured one week ago (A_0, Fig. 22 (c)) on a 12×12 image plane.

Impact of occlusion. We further conduct experiments to evaluate the imaging performance under various occlusions. We cover the image grid with common materials, such as cloth, opaque plastic sheets, and paper covers. As shown in Fig. 23, our system can correctly reconstruct the objects while they are visually obstructed, making it possible for our system to perform in visually obstructed scenarios, such as security checks and product inspections.

Impact of environment. We evaluate the imaging performance in various real-world environments, including an open indoor area, a narrow corridor, and a clustered room. As shown in Fig 24, the RMSE offsets of the corridor and the cluster room scenarios compared to the open space scenario are 2.44% and 15.54%, respectively. This indicates that MIMSID is resilient to environmental changes. We believe this is due to the introduction of range-gating, which can effectively filter out the out-of-ROI reflections and multipaths.

8.9 Real-object Imaging

As shown in Sec. 8.8, MIMSID can generate high-quality images of occluded subjects, well-suited for applications such as security checks and assembly line quality inspections. Fig. 25 simulates two typical real-world use cases of MIMSID, metal weapon detections and product defect detections. In the first example, the images of a gun, a cutter knife, and a metallic shovel are accurately reconstructed, indicating the effectiveness of MIMSID in metal weapon detections. In the second example, disc shapes are generated with different defect conditions, proving MIMSID is capable of detecting defective products with broken shapes.

Real-world objects may involve movements. Fig. 25(c) simulates an example of object movements on an assembly line. Our system is able to avoid motion blur with an object speed of $1.44m/s$ [36] owing to the short framing rate of our system (*i.e.*, $0.24ms$ per frame). The objects can be treated as static in each frame, especially when the object is moving relatively slowly and uniformly, such as in assembly lines and through security inspections. The objects can be seen as sequential images with different positions and achieve a median RMSE of 0.074.

The imaging scenario can be further extended to 3D imaging with multiple image planes at different distances. Fig. 26 shows a 3D image of a gun and a cutter knife with 4 image planes. The image planes are separated by $5mm$ and construct a $20cm \times 20cm \times 2cm$ 3D imaging space $30cm$ away from the transceiver. The shapes and positions of the objects are reconstructed clearly with an RMSE of 0.141, indicating the effectiveness in performing 3D imaging.

9 Limitation and Discussion

Reflecting surface conditions. Although MIMSID enhances the system's tolerance to varying surfaces and materials by introducing noise to the training samples, the reflectance properties of the object remain a crucial factor influencing imaging quality. Low-reflective surfaces, such as human body parts, reflect significantly less signal compared to metals, leading to noise-dominated received signals and channel mismatches between the object and the calibration matrix. Additionally, objects with mirror-like reflections pose challenges for imaging, as a substantial portion of the reflective signals is not captured by the receiver. We will continue to take measurements to develop more robust models that account for materials with diverse reflectance characteristics.

Recalibration cost. While MIMSID is resilient to environmental changes, the end-to-end optimization and passive metasurface design are dependent on the distance, size, and resolution of the image plane. The channel matrix A needs to be recalibrated if the image plane changes. To maintain the target within the predefined image plane at all times, one potential approach is to calibrate a large 3D image plane that encompasses all possible target locations. However, this approach significantly increases the calibration effort. By investigating automated calibration techniques, such as the use of robotic arms or rails, the time and labor costs associated with calibration could be reduced. This aspect is left for future exploration.

Pre-defined static metasurface. MIMSID employs a PCB-printed metasurface for virtual phase array expansion, which may limit the system's flexibility in adapting to new predefined image planes. One potential solution is to optimize the metasurface for a large 3D image plane that covers all possible target locations. However, this could significantly increase computational complexity. To facilitate rapid updates, simplify manufacturing, and reduce costs, paper-printed metasurfaces [28] could be considered as a replacement for the current metasurfaces. The phase manipulation capabilities of paper-printed metasurfaces at W-band require further investigation.

System misalignment. As MIMSID necessitates channel modeling between the transceiver, metasurface, and the image plane, precise alignment of these components is crucial for optimal performance. The predefined image plane is conceptual and does not require manual alignment. However, misalignment between the transceiver and the metasurface can adversely affect system performance. The lightweight metasurface is easy to deploy and align with the transceiver, and the entire system can be encapsulated in a protective enclosure to stabilize the relative positions of the components, thereby minimizing the risk of misalignment.

10 Conclusion

In this paper, we develop MIMSID, a novel system for high-resolution W-band mmWave imaging. Our system utilizes a small, COTS transceiver array without any mechanical movement. We design a passive transmissive metasurface at $77-81GHz$ with a high transmission rate and a wide range of phase manipulation to significantly improve the imaging resolution. In addition, we repurpose the state-of-the-art image generation diffusion models for image reconstruction and design a signal-to-image diffusion network for mmWave imaging. Our imaging system achieves a median RMSE of 0.061 for a $20cm \times 20cm$ image plane with $1cm$ resolution, indicating the effectiveness of MIMSID. We believe this work inspires further development of compact mmWave imaging systems.

Acknowledgments

We are grateful to all anonymous reviewers for their constructive comments and Yitong Fan's instruction on 3D-model designs. This work is supported in part by National Natural Science Foundation of China (No. 61936015), Natural Science Foundation of Shanghai (No. 24ZR1430600) and Shanghai Key Laboratory of Trusted Data Circulation and Governance, and Web3.

References

[1] M. Bayati and A. Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Transactions on Information Theory*, 57(2):764–785, 2011.

[2] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

[3] G. Cardoso, Y. J. E. Idrissi, S. L. Corff, and E. Moulines. Monte carlo guided diffusion for bayesian linear inverse problems. *arXiv preprint arXiv:2308.07983*, 2023.

[4] L. Chen, B. Yu, Y. Fu, J. Ren, H. Pan, J. Gummeson, and Y. Zhang. Pushing wireless charging from station to travel. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 46–61, 2024.

[5] G. Chi, Z. Yang, C. Wu, J. Xu, Y. Gao, Y. Liu, and T. X. Han. Rf-diffusion: Radio signal generation via time-frequency diffusion. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 77–92, 2024.

[6] D.-M. Chian, C.-K. Wen, C.-H. Wu, F.-K. Wang, and K.-K. Wong. A novel channel model for reconfigurable intelligent surfaces with consideration of polarization and switch impairments. *IEEE Transactions on Antennas and Propagation*, 2024.

[7] K. W. Cho, M. H. Mazaheri, J. Gummeson, O. Abari, and K. Jamieson. mmwall: A reconfigurable metamaterial surface for mmwave networks. In *Proceedings of the 22nd International Workshop on Mobile Computing Systems and Applications*, pages 119–125, 2021.

[8] F.-A. Croitoru, V. Hondru, R. T. Ionescu, and M. Shah. Diffusion models in vision: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(9):10850–10869, 2023.

[9] P. Dhariwal and A. Nichol. Diffusion models beat gans on image synthesis. *Advances in neural information processing systems*, 34:8780–8794, 2021.

[10] D. L. Donoho, A. Maleki, and A. Montanari. Message-passing algorithms for compressed sensing. *Proceedings of the National Academy of Sciences*, 106(45):18914–18919, 2009.

[11] C. Feng, X. Li, Y. Zhang, X. Wang, L. Chang, F. Wang, X. Zhang, and X. Chen. Rflens: metasurface-enabled beamforming for iot communication and sensing. In *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking*, pages 587–600, 2021.

[12] Y. Fu, Y. Zhang, Y. Lu, L. Qiu, Y.-C. Chen, Y. Wang, M. Wang, Y. Li, J. Ren, and Y. Zhang. Adaptive metasurface-based acoustic imaging using joint optimization. In *Proceedings of the 22nd Annual International Conference on Mobile Systems, Applications and Services*, pages 492–504, 2024.

[13] Y. Fu, Y. Zhang, H. Pan, Y. Lu, X. Li, L. Chen, J. Ren, X. Li, X. Zhang, and Y. Zhang. Pushing the limits of acoustic spatial perception via incident angle encoding. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 8(2):1–28, 2024.

[14] X. Gao, S. Roy, and G. Xing. Mimo-sar: A hierarchical high-resolution imaging algorithm for mmwave fmcw radar in autonomous driving. *IEEE Transactions on Vehicular Technology*, 70(8):7322–7334, 2021.

[15] J. Gollub, O. Yurduseven, K. P. Trofatter, D. Arnitz, M. F. Imani, T. Sleasman, M. Boyarsky, A. Rose, A. Pedross-Engel, H. Odabasi, et al. Large metasurface aperture for millimeter wave computational imaging at the human-scale. *Scientific reports*, 7(1):42650, 2017.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.

[17] J. Guan, S. Madani, S. Jog, S. Gupta, and H. Hassanieh. Through fog high-resolution imaging using millimeter wave radar. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11464–11473, 2020.

[18] J. Han, L. Li, S. Tian, G. Liu, H. Liu, and Y. Shi. Millimeter-wave imaging using 1-bit programmable metasurface: Simulation model, design, and experiment. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*, 10(1):52–61, 2020.

[19] J. Hu, H. Zhang, B. Di, L. Li, K. Bian, L. Song, Y. Li, Z. Han, and H. V. Poor. Reconfigurable intelligent surface based rf sensing: Design, optimization, and implementation. *IEEE Journal on Selected Areas in Communications*, 38(11):2700–2716, 2020.

[20] Q. Hu, J. Zhao, K. Chen, K. Qu, W. Yang, J. Zhao, T. Jiang, and Y. Feng. An intelligent programmable omni-metasurface. *Laser & Photonics Reviews*, 16(6):2100718, 2022.

[21] M. A. Iqbal, A. Anghel, M. Datcu, I. Ederra, and J. C. Iriarte. Assessment of mm-wave high resolution inverse sar imaging both with compact and sparse data. In *2023 20th European Radar Conference (EuRAD)*, pages 266–269. IEEE, 2023.

[22] A. Kazerouni, E. K. Aghdam, M. Heidari, R. Azad, M. Fayyaz, I. Hacihaliloglu, and D. Merhof. Diffusion models in medical imaging: A comprehensive survey. *Medical image analysis*, 88:102846, 2023.

[23] A. Kosuge, S. Suehiro, M. Hamada, and T. Kuroda. mmwave-yolo: A mmwave imaging radar-based real-time multiclass object recognition system for adas applications. *IEEE Transactions on Instrumentation and Measurement*, 71:1–10, 2022.

[24] H. Lai, G. Luo, Y. Liu, and M. Zhao. Enabling visual recognition at radio frequency. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 388–403, 2024.

[25] W. Lee, S. Jo, K. Lee, H. S. Park, J. Yang, H. Y. Hong, C. Park, S. K. Hong, and H. Lee. Single-layer phase gradient mmwave metasurface for incident angle independent focusing. *Scientific reports*, 11(1):12671, 2021.

[26] I. Loshchilov and F. Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.

[27] R. Ma and W. Hu. Rf-mediator: Tuning medium interfaces with flexible metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 155–169, 2024.

[28] R. Ma, S. Zheng, H. Pan, L. Qiu, X. Chen, L. Liu, Y. Liu, W. Hu, and J. Ren. Automs: Automated service for mmwave coverage optimization using low-cost metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 62–76, 2024.

[29] X. Meng and Y. Kabashima. Diffusion model based posterior sampling for noisy linear inverse problems. *arXiv preprint arXiv:2211.12343*, 2022.

[30] A. Mirbeik, R. Ashinoff, T. Jong, A. Aued, and N. Tavassolian. Real-time high-resolution millimeter-wave imaging for in-vivo skin cancer diagnosis. *Scientific Reports*, 12(1):1–10, 2022.

[31] A. Montanari, Y. Eldar, and G. Kutyniok. Graphical models concepts in compressed sensing. *Compressed Sensing*, pages 394–438, 2012.

[32] A. E. Olk, P. E. Macchi, and D. A. Powell. High-efficiency refracting millimeter-wave metasurfaces. *IEEE Transactions on Antennas and Propagation*, 68(7):5453–5462, 2020.

[33] H. Pan and L. Qiu. Passive metasurface-based low earth orbit ground station design. *Tsinghua Science and Technology*, 30(1):148–160, 2024.

[34] H. Pan and L. Qiu. Passive metasurface for interacting with electromagnetic signals, Sept. 19 2024. US Patent App. 18/608,421.

[35] H. Pan, L. Qiu, B. Ouyang, S. Zheng, Y. Zhang, Y.-C. Chen, and G. Xue. Pmsat: Optimizing passive metasurface for low earth orbit satellite communication. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023.

[36] A. Pedross-Engel, C. M. Watts, and M. S. Reynolds. High-throughput 3-d millimeter-wave imaging of packaged goods. In *2020 IEEE Radar Conference (RadarConf20)*, pages 1–6. IEEE, 2020.

[37] A. Prabhakara, V. Singh, S. Kumar, and A. Rowe. Osprey: A mmwave approach to tire wear sensing. In *Proceedings of the 18th international conference on mobile systems, applications, and services*, pages 28–41, 2020.

[38] L. Qiu, H. Pan, M. Ruichun, and S. Zheng. Metasurface design and placement, Jan. 2 2025. US Patent App. 18/757,113.

[39] A. Radford, J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[40] H. Regmi, M. S. Saadat, S. Sur, and S. Nelakuditi. Squigglemilli: Approximating sar imaging on mobile millimeter-wave devices. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(3):1–26, 2021.

[41] RNA Automation Limited. Advantages of automated rotary assembly line, 2024.

[42] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10684–10695, 2022.

[43] J. M. Schellberg, H. Regmi, and S. Sur. Mmsight: Towards robust millimeter-wave imaging on handheld devices. In *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, pages 117–126. IEEE, 2023.

[44] Z. Shaikhanov, S. Badran, H. Guerboukha, J. Jornet, D. Mittleman, and E. Knightly. Metafly: Wireless backhaul interception via aerial wavefront manipulation. In *2024 IEEE Symposium on Security and Privacy (SP)*, pages 151–151. IEEE Computer Society, 2024.

[45] D. M. Sheen, D. L. McMakin, and T. E. Hall. Three-dimensional millimeter-wave imaging for concealed weapon detection. *IEEE Transactions on microwave theory and techniques*, 49(9):1581–1592, 2001.

[46] J. A. Solano-Pérez, J.-M. Molina-Garcia-Pardo, J.-V. Rodríguez, L. Jofre, and A. Mateo-Aroca. Comparison of using mm-wave fmcw radar if signals against frequency-domain vnas in the application of the multifrequency bifocusing (mfbf) imaging algorithm. *AEU - International Journal of Electronics and Communications*, April 2020.

[47] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. *arXiv preprint arXiv:2010.02502*, 2020.

[48] Y. Song and S. Ermon. Generative modeling by estimating gradients of the data distribution. *Advances in neural information processing systems*, 32, 2019.

[49] Y. Song, C. Ge, L. Qiu, and Y. Zhang. 2ace: Spectral profile-driven multi-resolutional compressive sensing for mmwave channel estimation. In *Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages

41–50, 2023.

[50] Y. Song, H. Pan, L. Ge, L. Qiu, S. Kumar, and Y.-C. Chen. Microsurf: Guiding energy distribution inside microwave oven with metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1346–1360, 2024.

[51] X. Tan, Z. Sun, D. Koutsonikolas, and J. M. Jornet. Enabling indoor mobile millimeter-wave networks based on smart reflect-arrays. In *IEEE INFOCOM 2018-IEEE Conference on Computer Communications*, pages 270–278. IEEE, 2018.

[52] W. Tang, J. Y. Dai, M. Z. Chen, K.-K. Wong, X. Li, X. Zhao, S. Jin, Q. Cheng, and T. J. Cui. Mimo transmission through reconfigurable intelligent surface: System design, analysis, and implementation. *IEEE journal on selected areas in communications*, 38(11):2683–2699, 2020.

[53] S. Tariq, S. I. Naqvi, N. Hussain, and Y. Amin. A metasurface-based mimo antenna for 5g millimeter-wave applications. *IEEE Access*, 9:51805–51817, 2021.

[54] Texas Instruments. Cascade coherency and phase shifter calibration application note, 2020.

[55] Texas Instruments. Imaging radar using cascaded mmwave sensor reference design, 2020.

[56] Q. Wang, E. T. Rogers, B. Gholipour, C.-M. Wang, G. Yuan, J. Teng, and N. I. Zheludev. Optically reconfigurable metasurfaces and photonic devices based on phase change materials. *Nature photonics*, 10(1):60–65, 2016.

[57] Y. Wang, H. Pan, L. Qiu, L. Zhong, J. Liu, R. Ma, Y.-C. Chen, G. Xue, and J. Ren. Gpms: Enabling indoor gnss positioning using passive metasurfaces. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1424–1438, 2024.

[58] Y. Wen, S. Wei, J. Wei, J. Liang, X. Zhang, and J. Shi. Non-line-of-sight imaging of hidden moving target using millimeter-wave inverse synthetic aperture radar. In *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, pages 555–558. IEEE, 2022.

[59] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.

[60] M. E. Yanik, D. Wang, and M. Torlak. Development and demonstration of mimo-sar mmwave imaging testbeds. *IEEE Access*, 8:126019–126038, 2020.

[61] W. Yu, H. He, X. Yu, S. Song, J. Zhang, R. Murch, and K. B. Letaief. Bayes-optimal unsupervised learning for channel estimation in near-field holographic mimo. *IEEE Journal of Selected Topics in Signal Processing*, 2024.

[62] F. Zhang, C. Wu, B. Wang, and K. R. Liu. mmeye: Super-resolution millimeter wave imaging. *IEEE Internet of Things Journal*, 8(8):6995–7008, 2020.