One Walk is All You Need: Data-Efficient 3D RF Scene Reconstruction with Human Movements

Yiheng Bian[1] Zechen Li[1] Lanqing Yang[1*] Hao Pan[1] Yezhou Wang[1] Longyuan Ge[1]
Jeffery Wu[1] Ruiheng Liu[1] Yongjian Fu[2] Yichao Chen[1] Guangtao Xue[1]

[1]Shanghai Jiao Tong University [2]Central South University

{byhbye123, yanglanqing, zechlee, panhao09, yezhouwang, gly2000,
jeffery2019, liutuiheng, yichao, gt_xue}@sjtu.edu.cn, fuyongjian@csu.edu.cn

Abstract

Reconstructing 3D Radiance Field (RF) scenes through opaque obstacles is a long-standing goal, yet it is fundamentally constrained by a laborious data acquisition process requiring thousands of static measurements, which treats human motion as noise to be filtered. This work introduces a new paradigm with a core objective: to perform fast, data-efficient, and high-fidelity RF reconstruction of occluded 3D static scenes, using only a single, brief human walk. We argue that this unstructured motion is ***not noise****, but is in fact an* ***information-rich signal*** *available for reconstruction. To achieve this, we design a factorization framework based on composite 3D Gaussian Splatting (3DGS) that learns to model the dynamic effects of human motion from the persistent static scene geometry within a raw RF stream. Trained on just a single 60-second casual walk, our model reconstructs the full static scene with a Structural Similarity Index (SSIM) of* ***0.96****, remarkably outperforming heavily-sampled state-of-the-art (SOTA) by* ***12%****. By transforming the human movements into its valuable signals, our method eliminates the data acquisition bottleneck and paves the way for on-the-fly 3D RF mapping of unseen environments.*

1. Introduction

The unique ability of Radiance Field (RF) signals to penetrate opaque obstacles offers a new frontier for building high-fidelity 3D models of non-line-of-sight (NLOS) spaces, with critical applications in robotics, augmented reality, and smart infrastructure [1, 35].

However, this promising frontier has been fundamentally constrained by a bottleneck: a costly, brute-force approach to data acquisition. Whether using early tomographic techniques [1] or modern neural representations like Neural Radiance Fields (NeRF) [22, 26] and 3D Gaussian Splatting (3DGS) [40, 43], all existing methods are slaves to a cripplingly laborious process. They rely on robotic platforms to scan dense grids over hours, capturing thousands of calibrated static measurements to disambiguate the scene's geometry. This insatiable demand for data has relegated high-fidelity RF mapping to the laboratory, making rapid, real-world deployment an impossibility.

The field's foundation rests on a seemingly indisputable dogma: the reconstruction of static scenes must rely on **static measurements**, while **dynamic events**, especially unpredictable human motion, are a source of chaotic noise that must be aggressively filtered or entirely avoided [33].

In this paper, we fundamentally challenge this long-standing dogma and posit the contrary: the chaotic interference from unstructured human motion is not a source of noise to be mitigated, but rather an information-rich signal for reconstructing the static world. A person walking through a room is not an obstacle to be mitigated, but a powerful, active probe. Each step casts a unique "RF shadow", creating a cascade of complex diffractions and occlusions. This single, continuous dynamic event implicitly scans the environment from **thousands of virtual viewpoints**, revealing geometric details of the static background that a sparse grid of static measurements could never see. **The "noise" is the solution.**

Our argument begins with a new theoretical foundation. We provide an analysis demonstrating why a single, casual one-minute walk can provide geometric information equivalent to deploying a dense array of up to 10x more physical RF sensors. Critically, we then demonstrate that this rich, dynamic information can only be effectively harnessed by models exhibiting **linear superposition**. The complex, additive nature of multi-path scattering from a static background and a dynamic human necessitates a model that can linearly combine these components without destructive interference.

This principle leads us directly to 3D Gaussian Splatting (3DGS), whose rendering process, a linear

*Corresponding author.

summation of Gaussian contributions, is ideally suited for this task. Based on this insight, we propose a principled disentanglement framework. Instead of training a single monolithic model, we employ a **two-stage strategy**. First, we train a dense background 3DGS model on a minimal static dataset to represent the background environment. Second, we introduce a new, sparse set of "human Gaussians" and train them exclusively on the dynamic RF stream together with frozen background model , allowing them to capture the human-induced perturbations. The final, high-fidelity scene is rendered by simply adding these two linear models together, a process that inherently preserves the integrity of the static background while incorporating the rich geometric cues from the human motion.

Experiments on 3 different scenes show that our method significantly outperforms state-of-the-art (SOTA) baselines. More tellingly, we show that all SOTA methods improve when trained on the human-present data using our framework, proving the intrinsic value of the motion-induced signal. Our extensive ablations reveal crucial insights: (i) The performance gain does not stem from merely increased data volume, but from the **structured information within the perturbations**. (ii) Naive **end-to-end approach fails**, as it catastrophically overfits to transient noise; our two-stage framework is essential to isolate the signal. (iii) The primary **benefit arises from rich, diffuse scattering**, not simple specular reflections, and is most effective when the person's path intersects the transceivers in compact spaces.

Our contributions can be summarized as follows:

- We are the first to propose a new paradigm for RF reconstruction that repurposes unstructured human motion from debilitating noise into the primary signal for modeling the static environment.
- We propose a principled framework based on the linear superposition property of 3DGS, enabling the effective disentanglement of static and dynamic scene components.
- Experiments show that we achieve Structure Similarity Index Measure (SSIM) of up to 0.96, surpassing heavily-sampled static SOTA methods by 12%, showing that a single 60-second walk can replace massive sensor arrays, thereby eliminating the data acquisition bottleneck.

2. Related Work

Traditional RF Scene Reconstruction. Modeling the physical world with RF signals is a fundamentally challenging, ill-posed problem due to the complex nature of multi-path propagation [3, 17, 18, 39]. Consequently, the dominant paradigm for high-fidelity reconstruction has been to overcome this ambiguity through dense spatial sampling. Early works in computational imaging used techniques akin to tomography, requiring extensive measurements to form a coherent image [1]. Even modern methods, which aim to reconstruct detailed channel information like the power angular spectrum (PAS), are bound by this constraint [20, 25, 27]. They typically rely on robotic platforms to meticulously scan an environment, capturing thousands of data points from a dense grid of known locations to solve the complex inverse problem.

Neural Representations for RF Scene Modeling. The rise of neural fields has offered a powerful new toolset for this task. Inspired by the success of Neural Radiance Fields (NeRF)[26] in computer vision, researchers have adapted these techniques to the RF domain. Works like NeRF2[44], NeWRF [22], and WiNeRT [31] have demonstrated the ability to learn a continuous representation of an RF scene from discrete samples, enabling interpolation of channel properties at unmeasured locations. More recently, following the trend in vision, methods have shifted towards 3D Gaussian Splatting (3DGS)[11–13, 19, 21, 29]for its superior training speed and rendering efficiency. WRF-GS[40] and RF-3DGS [43] have successfully replaced the MLP backbone of RF NeRFs with 3D Gaussians, achieving comparable or better quality with significantly reduced computational cost. However, a critical thread unites all these advanced methods: they are designed exclusively for static scenes and are predicated on the availability of the same dense, static, and painstakingly collected datasets.

Handling Dynamics in Scene Representation. The concept of dynamics has been treated in two starkly different ways by the vision and RF communities. In computer vision, dynamic scenes are a primary object of study. Methods like D-NeRF [32] and Dynamic 3D Gaussians [23] explicitly model the motion of objects and people with the goal of reconstructing and rendering the dynamic elements themselves.

Researchers have worked to isolate the minute signal variations from breathing and heartbeats by aggressively filtering out the larger interference from body motion [33], or have used the macro-level disturbances to classify activities or detect presence [9]. In all cases, the dynamic interference is either a nuisance to be removed or a low-resolution signal for a classification task, not a tool for high-fidelity reconstruction of the surrounding static world.

3. Background

3.1. Primer on Electromagnetics

Radiance field reconstruction requires modeling complex propagation behaviors such as reflection, transmission, refraction, diffraction, and absorption, alongside multipath effects. We quantify these behaviors using the following mathematical models.

3.2. Propagation Formulations

Path Loss. The attenuation of EM waves over a distance d is calculated as: $\text{pathloss} = 20\log_{10}\frac{4\pi df}{c}$,

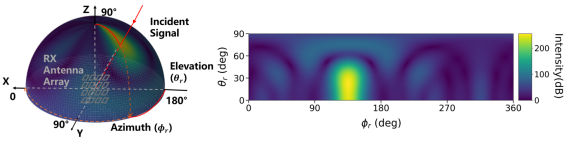


Figure 1. An example of the power angular spectrum.

where c is the speed of light and f is the frequency.

Reflection, Transmission, and Absorption. When waves encounter an obstacle, their behavior is governed by the material's permittivity (ϵ_m) and conductivity (μ_m). Using equivalent circuit models[16], we derive the reflection (R), transmission (T), and absorption (A) rates:

$$R = \left| \frac{\sqrt{\eta_m} - \sqrt{\eta_0}}{\sqrt{\eta_m} + \sqrt{\eta_0}} \right|^2 \tag{1}$$

$$T = \left| \frac{2\sqrt{\eta_0}}{\sqrt{\eta_m} + \sqrt{\eta_0}} \right|^2 \tag{2}$$

$$A = 1 - R - T \tag{3}$$

where $\eta_m = \frac{\mu_m}{\epsilon_m}$ and $\eta_0 = \frac{\mu_0}{\epsilon_0}$ are the impedances of the material and air.

The relationship between the angle of refraction (θ_t) and incidence (θ_i) is:

$$\frac{\sin(\theta_t)}{\sin(\theta_i)} = \sqrt{\frac{\epsilon_m \mu_m}{\epsilon_0 \mu_0}} \tag{4}$$

Diffraction. Edge diffraction is modeled using the Uniform Theory of Diffraction (UTD)[34]. The diffracted field E_d relates to the incident field E_i via:

$$E_d = E_i D(\theta_i, \theta_d, k, \epsilon_m, \mu_m) F \tag{5}$$

Here, $D(\cdot)$ is the diffraction coefficient, F is a polarization factor, and $k = \frac{2\pi f}{c}$ is the wavenumber.

3.3. Power Angular Spectrum (PAS)

To characterize the spatial distribution of the radiance field, we employ the Power Angular Spectrum (PAS). The PAS is represented as a 2D image of dimensions 90×360, where: the x-axis represents the **Azimuth** ($0^\circ \sim 360^\circ$); the y-axis represents the **Elevation** ($0^\circ \sim 90^\circ$). Each pixel value corresponds to the RF signal energy received from that direction, as illustrated in Fig. 1.

4. Preliminary:Why Human Motion Helps

In this section, we establish the physical foundation for our claim that human motion aids RF reconstruction. We will show that the human body acts as a mobile electromagnetic relay in section 4.1, prove that its weak scattered signals are sufficient for high-fidelity sensing in section 4.2, and demonstrate how motion provides a powerful form of spatial diversity crucial for reconstruction in section 4.3.

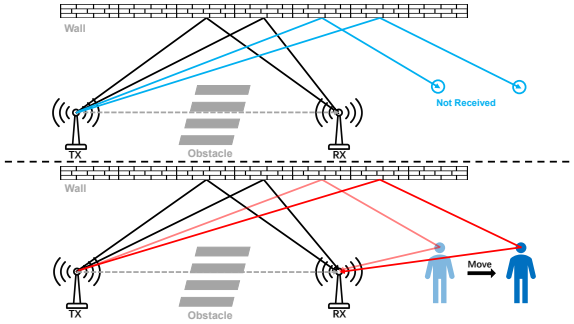


Figure 2. **Human motion creates new RF propagation paths.** The upper: no humans; the lower: with human movements. Although the body absorbs the signal, the scattered portion illuminates these otherwise invisible regions, and movement across multiple positions creates vast spatial diversity equivalent to thousands of virtual measurements.

4.1. Human Body as Electromagnetic Relay

When a transmitter and receiver are separated by obstacles, traditional RF sensing fails. However, a walking human—with high dielectric constant ($\epsilon_r \approx 40$) [7] acts as a *mobile electromagnetic relay*. Although the body absorbs 70% of incident energy, the scattered 30% is sufficient because it only needs to exceed the noise floor (∼-90 dBm), not compete with direct paths (∼-40 dBm). This 50 dB relaxation enables weak scattering to reveal occluded regions.

The human scattering effect consists of three primary components (a detailed proof is provided in **Appendix A**):

1. **Specular reflection** (∼10%): Mirror-like bounce from body surface, $\Gamma \approx |\frac{\sqrt{\epsilon_{\text{body}}}-1}{\sqrt{\epsilon_{\text{body}}}+1}|^2 \approx 0.47$ power coefficient, spatially averaged ≈10%.
2. **Diffuse scattering** (∼15%): Surface roughness (skin texture, clothing) creates Lambert-distributed scatter with radar cross-section $\sigma_{\text{RCS}} \approx 0.3$-$1.5$ m^2 at 2.4 GHz [10].
3. **Volume scattering** (∼5%): Internal tissue inhomogeneities (muscle, bone interfaces) cause multiple internal reflections before emerging.

The remaining 70% is absorbed as heat due to water content ($\epsilon''_{\text{body}} \approx 20$) [7], consistent with FCC SAR limits.

4.2. Why Weak Scattering Suffices?

Key insight: Scattered signals need only 10-15 dB SNR for reconstruction, achievable within 10 m despite ∼30 dB loss from body scattering. The radar equation for bistatic scattering shows:

$$P_{\text{scatter}} = P_t G_t G_r \lambda^2 \sigma_{\text{RCS}} / [(4\pi)^3 d_1^2 d_2^2] \tag{6}$$

where d_1, d_2 are TX-human and human-RX distances. With conservative parameters: $P_t = 20$ dBm, $\sigma_{\text{RCS}} = 0.3$ m^2, $\lambda = 0.125$ m, we achieve 15 dB SNR at 10 m total path, sufficient for phase-coherent measurements.

4.3. Spatial Diversity from Motion

As the human walks, their body samples different spatial positions, each providing unique scattering geometry. The effective information gain is:

$$\mathcal{I}_{\text{total}} = K \cdot N_{\text{pos}} \cdot (1 - \rho_{\text{corr}}) \tag{7}$$

where K is the effective rank of observations per position ($\approx$8 for our setup), N_{pos} is the number of positions sampled during the walk ($\approx$35 for 10-second walk), and ρ_{corr} is the spatial correlation ($\approx$0.2 for 30 cm spacing). This yields $\mathcal{I}_{\text{total}} \approx 224$ effective measurements—equivalent to deploying 224 static RF sensors. Fisher information analysis confirms this provides sufficient observability for 3D reconstruction of occluded regions.

The human body's $\epsilon_r \approx 40$ ensures strong perturbations to the field, making the separation tractable. Mathematical justification via perturbation analysis is provided in **Appendix B**.

4.4. Conclusion

In summary, we have shown that the human body acts as an effective, mobile RF relay whose scattered signals are sufficiently strong for sensing. Crucially, the spatial diversity gained from motion provides an information gain equivalent to a dense sensor array. This establishes a key principle: incorporating dynamic elements like a moving person is a powerful method to enrich the information content of RF datasets, leading to more robust and accurate field reconstruction.

5. Method Design

5.1. Overview

Modeling task: Existing 3DGS-based RF reconstruction has been limited to static scenes, such as those with a moving transmitter (TX) or receiver (RX). However, these frameworks are ill-equipped to handle dynamic environments involving human mobility. To address this limitation, we introduce a more challenging task: reconstructing the radiance field in the presence of a moving person. Our work leverages a newly **generated** dynamic dataset that correlates human positions with their impact on the PAS. The objective is to train a unified model capable of accurately reconstructing the PAS for **any given human position and any corresponding antenna position in both moving TX and RX tasks**.

Our Approach: We conceptualize the radiance field in a dynamic scene as a superposition of two distinct components: a static background field $F_{\text{bg}}(\mathbf{r})$ and a dynamic, human-induced perturbation field $\Delta F(\mathbf{r}, h)$. This decomposition is formally expressed as:

$$\mathbf{y}_{\text{total}}(\mathbf{r}, h) = \underbrace{F_{\text{bg}}(\mathbf{r})}_{\text{static background}} + \underbrace{\Delta F(\mathbf{r}, h)}_{\text{dynamic perturbation}} \tag{8}$$

This decomposition enables a two-stage training strategy. First, we train a background model on a static (human-absent) dataset. Second, we freeze this pre-trained model and train a dedicated perturbation model on the dynamic (human-present) dataset. This isolates the optimization to solely capture the human-induced effects, leading to an efficient and accurate reconstruction of the complete dynamic field.

5.2. Theoretical Foundation: Two-Stage Training Rationale

Building upon the physical principle established in Section 3, we now present the theoretical foundation for our two-stage training strategy when incorporating human body dynamics.

In this section, we elucidate the rationale behind our methodological design by answering the following three fundamental questions.

5.2.1. Why Two-Stage Training?

In this section, we explain why a two-stage training strategy is necessary. From Eq. 8, the measurement model with human presence can be expressed as:

$$\mathbf{y}(h) = (\mathbf{\Phi}_{\text{bg}} + \Delta\mathbf{\Phi}(h))\mathbf{x} + \mathbf{n} \tag{9}$$

Jointly optimizing both $\mathbf{\Phi}_{\text{bg}}$ and $\Delta\mathbf{\Phi}(h)$ leads to:

$$\min_{\mathbf{x}, \mathbf{\Phi}_{\text{bg}}, \Delta\mathbf{\Phi}} \sum_{h \in \mathcal{H}} ||\mathbf{y}(h) - (\mathbf{\Phi}_{\text{bg}} + \Delta\mathbf{\Phi}(h))\mathbf{x}||^2 \tag{10}$$

However, this approach suffers from **background-perturbation coupling**, where gradients contaminate each other.

Lemma 1 (Gradient Cleanness via Separation)**.** *If $\mathbf{\Phi}_{bg}$ is estimated first using human-free data:*

$$\nabla_{\mathbf{\Phi}_{bg}} \mathcal{L}_{stage1} = -2(\mathbf{y}_{static} - \mathbf{\Phi}_{bg}\hat{\mathbf{x}})\hat{\mathbf{x}}^T \tag{11}$$

this gradient contains no $\Delta\mathbf{\Phi}$ contamination, leading to more stable convergence.

Subsequently, freezing $\mathbf{\Phi}_{bg}$ in Stage 2:

$$\nabla_{\Delta\mathbf{\Phi}} \mathcal{L}_{stage2} = -2(\mathbf{y}(h) - \mathbf{\Phi}_{bg}\hat{\mathbf{x}} - \Delta\mathbf{\Phi}(h)\hat{\mathbf{x}})\hat{\mathbf{x}}^T \tag{12}$$

focuses gradient flow exclusively on the residual perturbation without corrupting the learned background.

Physical interpretation: The two-stage strategy mirrors the signal decomposition in Eq. 8—first model the static environment, then capture human-induced variations. See proof in **Appendix C**.

5.2.2. Why 3D Gaussian Splatting?

We adopt a 3D Gaussian Splatting (3DGS) framework to model the radiance field, as its representation with discrete Gaussians aligns well with the Huygens-Fresnel principle of secondary wave sources [24]. This choice is further supported by the high efficiency and fidelity demonstrated in recent works like WRF-GS [40] and GSRF [42]. Each Gaussian models local

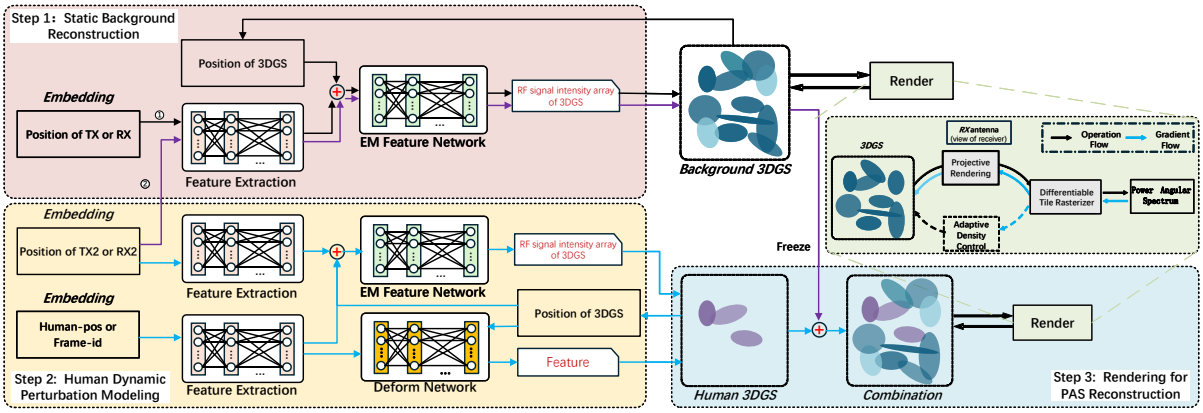


Figure 3. Architecture of the system.

EM propagation via spatial attributes, an attenuation factor $\delta(G_j)$ for path occlusion, and a signal radiance $Sig(G_i)$ representing emitted energy. The signal is rendered through volumetric accumulation:

$$I(p) = \sum_{i=1}^{N} \left(\prod_{j=1}^{i-1} \delta(G_j) \right) Sig(G_i) \qquad (13)$$

where N is the number of Gaussians covering the pixel p, and $\delta(G_j)$ is the attenuation (or transmission) factor of the j-th Gaussian in the sorted sequence.

The linearity of this rendering equation is crucial for our two-stage strategy. A dynamic scene is rendered by the direct superposition of new "human" Gaussians onto a pre-trained, frozen set of background Gaussians. Since the background set acts as a constant baseline, the optimization can focus on the dynamic, human-induced perturbations without conflicting gradients. Given this inherent compatibility, we select the state-of-the-art WRF-GSplus as our base model.

5.2.3. Why a Sparse Human Representation?

We explain how a sparse set of new Gaussians can accurately model complex human-induced perturbations. We explain this by modeling the human-induced perturbation as a low-rank phenomenon, where the complex field changes can be decomposed into a small set of principal scattering modes. This allows us to use a lightweight network to capture the dynamic effects without over-parameterization.

Lemma 2 (Low-Rank Scattering Model)**.** *The human-induced perturbation exhibits low-rank structure:*

$$\Delta F(\mathbf{r}, h) \approx \sum_{k=1}^{K} \alpha_k(h)\phi_k(\mathbf{r}), \quad K \ll M \qquad (14)$$

where K principal modes suffice to capture scattering effects.

The intuition for this low-rank property, with a formal derivation in **Appendix D**, stems from three physical arguments:

1. **Limited Spatial Extent:** A human occupies a localized volume, constraining its RF impact (via scattering and shadowing) and thus preventing global complexity.
2. **Dominant Scattering Directions:** Energy scattered from the human body is anisotropic, concentrating along a few dominant paths (e.g., specular components), which limits the number of basis modes required.
3. **Analogy to Signal Processing:** This is analogous to the compact representation of moving targets in other domains, such as Doppler spectrum compression in Synthetic Aperture Radar (SAR).

Implementation implication: Use a sparse set of Gaussian primitives, governed by a lightweight **deformable network**, to effectively model $\Delta\mathbf{\Phi}(h)$ and avoid over-parameterization.

5.3. Our Approach

Our model is founded on the 3DGS technology and employs a **two-stage training strategy** to effectively model dynamic scenes. This strategy first learns the static background environment and then sequentially models the perturbations caused by human movement. Crucially, the second stage integrates a human mobility embedding module to explicitly learn the complex relationship between the person's position and the resulting impact on EM wave propagation. The overall architecture is illustrated in Figure 3.

Stage 1: Static Background Reconstruction. The first stage focuses on reconstructing the static radiation field. We use the static (human-absent) dataset to train a background model for either the moving RX or moving TX task. This model consists of two main components: a standard 3DGS model for PAS synthesis, and an EM feature network that learns the propagation-related parameters of the 3D Gaussians. The positions of the 3D Gaussians are initialized from the scene's point cloud, while other properties are randomly initialized or set to default values. The EM feature network takes the positions of the mobile antenna (either TX or RX) and the 3D Gaussians as input, and outputs the signal radiance for each Gaussian

$Sig(G_{bg})$. The EM feature network can be expressed as follows:

$$F_{\Theta bg} : (P(G_{bg}), P(antenna)) \rightarrow (Sig(G_{bg}))$$

Stage 2: Human Dynamic Perturbation Modeling. In the second stage, we freeze all parameters of the pre-trained background model. The background Gaussians remain part of the rendering process but are excluded from further optimization. We then introduce a new, sparsely and randomly initialized set of 3D Gaussians dedicated to modeling the human's influence. This new set of Gaussians is associated with two distinct networks:

- **EM Feature Network:** Similar to the background model, but with an additional input—the person's position—to predict the radiance of the new Gaussians $Sig(G_{man})$ $F_{\Theta man} : (P(G_{man}), P(antenna), P(man)) \rightarrow (Sig(G_{man}))$
- **Deformation Network:** To explicitly model how the person's location affects the geometry of the perturbation field, we introduce a deformation network. This network takes the positions of the new Gaussians and the human's location as input, and outputs their displacement, rotation, and scaling components.

$$D_{\Theta man} : (P(G_{man}), P(man)) \rightarrow (Features(G_{man})) \tag{15}$$

This two-network setup allows the model to learn a compact and dynamic representation of human-induced effects like scattering and shadowing.

Rendering and Optimization. In both stages above, the final PAS is rendered by splatting all active 3D Gaussians onto the RX view hemisphere using a tile-based differentiable rasterizer. During training, the loss gradient between the predicted PAS and the ground truth is backpropagated to update the trainable parameters—the properties of the Gaussians and the weights of the associated neural networks. As noted, in Stage 2, only the parameters of the newly introduced "human" Gaussians and their corresponding EM feature and deformation networks are updated.

Stage 3: Power Angular Spectrum Rendering. To render the Power Angular Spectrum (PAS) at a given receiver (RX) location, we first project the 3D Gaussians onto the RX's 2D image plane. This projection is achieved by converting each Gaussian's 3D position into spherical coordinates (azimuth and zenith angles) relative to the RX's local frame, and then scaling these angles to the corresponding pixel coordinates on the PAS image. Following this transformation, the final PAS image is synthesized using a differentiable tile-based rasterizer. Within this process, the projected 2D Gaussians are sorted by depth, and the pixel value $I(p)$ is computed by accumulating the signal of each Gaussian, $Sig(G_i)$, attenuated by the cumulative product of the attenuation factors, $\delta(G_j)$, of all Gaussians positioned in front of it. The formulation of rendering is shown in equation 13.

5.4. Model Training

We utilize the adaptive density control mechanism of 3DGS to dynamically refine the Gaussian set during training. This process periodically densifies Gaussians in under-reconstructed areas while pruning those with negligible contributions. This dual-action approach prevents both under-reconstruction and over-densification, ensuring an efficient allocation of the model's representational capacity.

The parameters of the 3D Gaussians, the EM feature network, and the deformation network are optimized via stochastic gradient descent (SGD)[5]. The objective is to minimize a composite loss function between the predicted PAS and the ground truth. The loss is a weighted sum of the L1 loss and the Structural Similarity Index Measure (SSIM) loss, where λ is set to 0.2:

$$\mathcal{L} = (1 - \lambda)\mathcal{L}_{\text{L1}} + \lambda\mathcal{L}_{\text{SSIM}} \tag{16}$$

6. Evaluation

6.1. Dataset and Experimental Setup

Dataset: We generate a simulated dataset for Power Angular Spectrum (PAS)[36] reconstruction in indoor environments with human mobility using the NVIDIA Sionna™ simulator[15]. The dataset includes two scenarios: one with a moving 4×4 Uniform Rectangular Array (URA) receiver (RX) and another with a moving omnidirectional transmitter (TX).As shown in **Appendix E** , we evaluate in three typical indoor scenes, in which we sample 900 mobile antenna positions. We use the Bartlett method[37] to generate ground truth PAS images. For each scene, we create two datasets: an *empty-scene* dataset with multiple static RX/TX locations to train the baseline model, and a *human-present* dataset where a person moves through 35 locations in 1 minute, creating perturbed spectrograms for the same static RX/TX positions.

Training and Evaluation: We employ a two-stage strategy: a 3DGS model is first trained on static data to learn the background field, then frozen. Subsequently, a sparse set of Gaussians is optimized on dynamic data to model human-induced perturbations. The dynamic data is split into 81% for training and 19% for validation and testing, featuring unseen human and antenna positions. For evaluation, we withhold 20% of the RX/TX locations as a **Test Set** to assess generalization to unseen positions, using the remaining 80% for training. Our primary evaluation metrics are the Mean and Median Structural Similarity Index Measure (**SSIM**)[38] at **Test Set**.

Compared Methods: We benchmark our proposed method (**Ours**) against two baselines: 1) **Baseline**, a modified WRFGS-Plus [41] model trained only on the empty-scene data, and 2) **End-to-End**, a single-stage model trained on the human-present data with human location as a direct input.

Table 1. Performance comparison on static (Human-absent) and dynamic (Human-present) datasets across three scenes.

Model	Dataset	Scene 1		Scene 2		Scene 3	
		Mean SSIM	Median SSIM	Mean SSIM	Median SSIM	Mean SSIM	Median SSIM
wrfgsplus	Human-absent	0.873	0.906	0.875	0.897	0.860	0.904
	Human-present	0.909	0.945	0.867	0.941	0.873	0.932
GSRF	Human-absent	0.758	0.758	0.781	0.781	0.814	0.814
	Human-present	0.818	0.818	0.826	0.826	0.851	0.851
wrfgsplus_mod	**Human-absent**	**0.844**	**0.886**	**0.847**	**0.897**	**0.830**	**0.891**
	Human-present	**0.895**	**0.937**	**0.918**	**0.954**	**0.852**	**0.924**

Table 2. Performance generalization across three different scenes for Moving Rx and Moving Tx (900 locations).

Scene	Method	Moving Rx		Moving Tx	
		Mean SSIM	Median SSIM	Mean SSIM	Median SSIM
Scene 1	Baseline	0.844	0.886	0.873	0.906
	End-to-End	0.884	0.929	0.881	0.924
	Ours (*man pos*)	**0.927**	**0.967**	**0.941**	**0.975**
Scene 2	Baseline	0.847	0.897	0.875	0.897
	End-to-End	0.868	0.894	0.877	0.937
	Ours (*man pos*)	**0.943**	**0.977**	**0.898**	**0.983**
Scene 3	Baseline	0.830	0.891	0.860	0.904
	End-to-End	0.821	0.882	0.856	0.916
	Ours (*man pos*)	**0.862**	**0.942**	**0.897**	**0.976**

Table 3. Baseline performance versus the number of Rx locations in an empty scene. While more measurement points behave better, the data acquisition cost becomes prohibitive.

Rx Locations	Mean SSIM	Median SSIM
900	0.844	0.886
1500	0.859	0.904
3600	0.911	0.947
6000	0.920	0.949
10000	0.929	0.952

6.2. Motivation: The Challenge of Dense Environmental Sampling

A conventional approach to improving reconstruction fidelity is to increase the density of measurements. We first quantify this effect by training our baseline model on the empty-scene dataset with a varying number of Rx locations. As shown in Table 3, increasing the Rx locations from 900 to 10,000 yields a significant improvement in Mean SSIM, from 0.844 to 0.929.

However, this highlights a critical trade-off: achieving high accuracy via dense sampling imposes a prohibitively high cost in terms of data acquisition time and effort in real-world scenarios. This limitation motivates our core research question: can we enhance reconstruction quality not by adding more static measurement points, but by efficiently leveraging dynamic, human-induced perturbations as a source of rich environmental information?

6.3. Main Results: Human Perturbation for Enhanced Reconstruction

Our method was evaluated across three distinct scenes for the Moving Rx task, where it consistently and substantially outperformed both the Baseline and the naive End-to-End approaches (detailed in Table 2). Focusing on Scene 1, our approach achieves a Mean SSIM of 0.927—a significant **+0.083 absolute improvement (9.8% relative gain)** over the strong Baseline (0.844). Notably, this performance surpasses that of a baseline trained with 6000 Rx locations, demonstrating that leveraging human motion can be more effective than a 6-fold increase in static measurements. The failure of the End-to-End method to achieve comparable gains further validates our two-stage strategy, which effectively isolates dynamic perturbations by first establishing a robust background model.

Fig. 4 presents partial visual results of PAS reconstruction. To illustrate the performance, we provide a comparison of spectrograms at six locations in Scene 1. We contrast the ground truth with the Baseline's prediction for the empty scene, alongside the measured data with our model's prediction for a dynamic scene where a person is present. It is clear that our method not only surpasses the current SOTA in terms of the SSIM, but also visually appears closer to the GT.

6.4. Generalization to Moving-Tx Task

We further assess the versatility of our method by applying it to the Moving-Tx task, where the transmitter's location is dynamic. The results in Table 2 confirm that our approach is task-agnostic. It again achieves a Mean SSIM of 0.938, significantly surpassing the Baseline (0.873). Notably, the End-to-End method's performance degrades below the baseline in this scenario, further highlighting the robustness of our decoupled learning strategy.

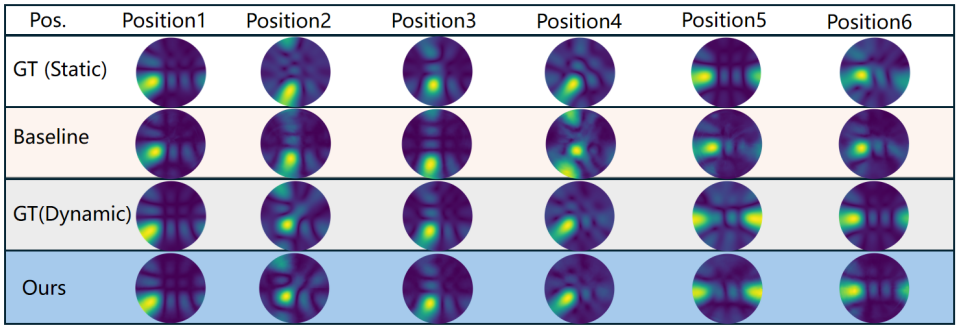


Figure 4. Overall PAS visualization at different positions.

Table 4. Ablation on the two-stage training strategy.

Method	Mean SSIM
End-to-End (on human data)	0.858
Continued Baseline (on empty data)	0.844
Ours (Two-Stage)	**0.927**

Table 5. Ablation on the role of perturbation data.

Method	Description	Mean SSIM
Baseline (Empty)	Trained on empty-scene data.	0.844
Baseline (Duplicated)	Trained on duplicated empty-scene data.	0.761
Baseline (Human)	Trained directly on human-present data.	0.895
Ours	**Modeling perturbations.**	**0.927**

Table 6. Model performance with different equivalent human materials.

Material	Mean SSIM (Test)	Median SSIM (Test)
Metal	0.915	0.958
Concrete	**0.935**	**0.972**
Human	0.927	0.967

6.5. Dynamic Data Augmentation Analysis

To validate the inherent benefit of the dynamic scene dataset, we trained the unmodified baseline models including the original WRF-GSplus, our adapted moving RX version, and GSRF-separately on the static (human-absent) and dynamic (human-present) datasets.

The results, detailed in Table 1, reveal a consistent and significant performance boost across all evaluated models when they are trained using the dynamic dataset. This improvement can be attributed to a form of **implicit physical regularization**. The subtle, physically-consistent perturbations introduced by the moving human act as a powerful data augmentation, compelling the models to learn a more robust and generalizable representation of the underlying EM propagation physics. This effect is particularly beneficial in mitigating overfitting, especially when the set of unique RX/TX locations is limited, leading to improved performance on the validation set.

6.6. Ablation Studies

Necessity of the Two-Stage Strategy. We first validate our two-stage training paradigm. As shown in Table 4, a single-stage End-to-End model (no two step test) fails to optimize the problem effectively. Furthermore, simply continuing to train the baseline on static data (continue test) yields no improvement, ruling out longer training times as the source of our gain. This confirms that our decoupled "background-first, perturbation-second" approach is critical for effective reconstruction.

The Intrinsic Role of Perturbation Information. We next verify that our performance gain comes from the rich information within perturbations, not just increased data volume. As shown in Table 5, naively increasing the dataset size by duplicating data causes severe overfitting and performance collapse (0.761). While using human-present data as a form of data augmentation provides a moderate boost (0.895), it is significantly outperformed by our method (0.927), which explicitly models the perturbations. This demonstrates that our success stems from interpreting perturbations as structured information rather than as generic augmented data.

6.7. Impact of Human Material Properties

We investigated how the perturbation source's material affects performance by simulating a human equivalent, metal, and concrete. As shown in Table 6, concrete achieved the best results. This is likely due to distinct scattering patterns: metal produces simple specular reflections, while the human-equivalent material's high dielectric loss causes significant signal absorption and low SNR. In contrast, concrete, as a low-loss dielectric, generates richer diffuse scattering, providing stronger and more effective constraints for the 3DGS inversion process.

7. Discussion and Conclusion

This work challenges a fundamental assumption in RF sensing: that human motion is noise to be filtered. We demonstrate the contrary—unstructured human motion is an information-rich signal available for reconstructing occluded static scenes. To this end, we introduced a new paradigm for 3D RF reconstruction, using a composite 3D Gaussian representation, successfully disentangles dynamic interference from a raw RF stream to produce a high-fidelity model of the static

scene. Experiments show our method achieves a 12% SSIM improvement over heavily-sampled baselines, using only a single 60-second walk.

Key challenges remain for future work. First, the framework can be extended to model multiple people simultaneously, which presents a complex data association problem. Second, we aim to achieve high-quality reconstruction using only temporal frame IDs without precise location data. Addressing these challenges will enable continuous, self-improving 3D mapping of crowded, real-world environments.

References

[1] Fadel Adib and Dina Katabi. See through walls with wi-fi! In *Proceedings of the ACM SIGCOMM 2013 conference on SIGCOMM*, pages 75–86, 2013. 1, 2

[2] Fadel Adib, Zach Kabelac, Dina Katabi, and Robert C Miller. 3d tracking via body radio reflections. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 317–329, 2014. 19

[3] Avishek Banerjee, Xingya Zhao, Vishnu Chhabra, Kannan Srinivasan, and Srinivasan Parthasarathy. Horcrux: Accurate cross band channel prediction. In *Proceedings of the 30th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2024. 2

[4] Dimitri P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 3rd edition, 2016. 14, 15, 16

[5] Léon Bottou. Stochastic gradient descent tricks. In *Neural networks: tricks of the trade: second edition*, pages 421–436. Springer, 2012. 6

[6] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, Cambridge, UK, 2004. 14, 16

[7] C.Gabriel. Compilation of the dielectric properties of body tissues at rf and microwave frequencies. Technical Report AL/OE-TR-1996-0037, U.S. Air Force Brooks Air Force Base, Occupational & Environmental Health Directorate, Radio-frequency Radiation Division, Brooks AFB, Texas, USA, 1996. Available from NTIS reference number AD-A309 764. 3, 11, 19

[8] KP Chethan, T Chakravarty, J Prabha, M Girish Chandra, and P Balamuralidhar. Polarization diversity improves rssi based location estimation for wireless sensor networks. In *2009 Applied Electromagnetics Conference (AEMC)*, pages 1–4. IEEE, 2009. 11

[9] Fangqiang Ding, Zhen Luo, Peijun Zhao, and Chris Xiaoxuan Lu. milliflow: Scene flow estimation on mmwave radar point cloud for human motion sensing. In *European Conference on Computer Vision*, pages 202–221. Springer, 2024. 2

[10] T. Dogaru, L. Nguyen, and C. Le. Computer models of the human body signature for sensing through the wall radar applications. *Army Research Laboratory Technical Report ARL-TR*, 4136:1–45, 2007. 3, 11, 19

[11] Z. Fan, K. Wang, K. Wen, et al. LightGaussian: Unbounded 3D Gaussian Compression with 15x Reduction and 200+ FPS. *Advances in Neural Information Processing Systems*, 37:140138–140158, 2024. 2

[12] G. Fang and B. Wang. Mini-Splatting: Representing Scenes with a Constrained Number of Gaussians. In *European Conference on Computer Vision*, pages 165–181, Cham, 2024. Springer Nature Switzerland.

[13] S. Girish, K. Gupta, and A. Shrivastava. Eagles: Efficient Accelerated 3D Gaussians with Lightweight Encodings. In *European Conference on Computer Vision*, pages 54–71, Cham, 2024. Springer Nature Switzerland. 2

[14] Roger A Horn and Charles R Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, UK, 2nd edition, 2012. 15

[15] Jakob Hoydis, Sebastian Cammerer, Fayçal Ait Aoudia, Avinash Vem, Nikolaus Binder, Guillermo Marcus, and Alexander Keller. Sionna: An open-source library for next-generation physical layer research. *arXiv preprint arXiv:2203.11854*, 2022. 6

[16] Xiaosong Hu, Shengbo Li, and Huei Peng. A comparative study of equivalent circuit models for li-ion batteries. *Journal of Power Sources*, 198:359–367, 2012. 3

[17] Wen Jiang, Boshu Lei, and Kostas Daniilidis. Fisherrf: Active view selection and uncertainty quantification for radiance fields using fisher information. *arXiv preprint arXiv:2311.17874*, 2023. 2

[18] Wei Jiang, Qiuheng Zhou, Jiguang He, Mohammad Asif Habibi, Sergiy Melnyk, Mohammed El-Absi, Bin Han, Marco Di Renzo, Hans Dieter Schotten, Fa-Long Luo, et al. Terahertz communications and sensing for 6g and beyond: A comprehensive review. *IEEE Communications Surveys & Tutorials*, 26 (4):2326–2381, 2024. 2

[19] Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 3d Gaussian splatting for real-time radiance field rendering. *ACM Transactions on Graphics*, 42(4):1–14, 2023. 2

[20] Robert G Kouyoumjian and Prabhakar H Pathak. A uniform geometrical theory of diffraction for an edge in a perfectly conducting surface. *Proceedings of the IEEE*, 62(11):1448–1461, 2005. 2

[21] J. C. Lee, D. Rho, X. Sun, et al. Compact 3D Gaussian Representation for Radiance Field. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 21719–21728. IEEE, 2024. 2

[22] Haofan Lu, Christopher Vattheuer, Baharan Mirzasoleiman, and Omid Abari. Newrf: A deep learning framework for wireless radiation field reconstruction and channel prediction. *arXiv preprint arXiv:2403.03241*, 2024. 1, 2

[23] Jonathon Luiten, Georgios Kopanas, Bastian Leibe, and Deva Ramanan. Dynamic 3d gaussians: Tracking by persistent dynamic view synthesis. *arXiv preprint arXiv:2308.09713*, 2023. 2

[24] JR Mahan, NQ Vinh, VX Ho, and NB Munir. Monte carlo ray-trace diffraction based on the huygens–fresnel principle. *Applied optics*, 57(18):D56–D62, 2018. 4

[25] James Clerk Maxwell. *A treatise on electricity and magnetism*. Clarendon press, 1873. 2

[26] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1): 99–106, 2021. 1, 2

[27] Han Na and Thomas F Eibert. A huygens' principle based ray tracing method for diffraction calculation. In *2022 16th European Conference on Antennas and Propagation (EuCAP)*, pages 1–4. IEEE, 2022. 2

[28] Yurii Nesterov. *Lectures on Convex Optimization*. Springer, 2nd edition, 2018. 14, 15, 16

[29] Michael Niemeyer, Fabian Manhardt, Marie-Julie Rakotosaona, Michael Oechsle, Daniel Duckworth, Rama Gosula, Keisuke Tateno, John Bates, Dominik Kaeser, and Federico Tombari. Radsplat: Radiance field-informed gaussian splatting for robust real-time rendering with 900+ fps. *arXiv preprint arXiv:2403.13806*, 2024. 2

[30] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer Science & Business Media, New York, NY, 2nd edition, 2006. 14, 15, 16

[31] Tribhuvanesh Orekondy, Pratik Kumar, Shreya Kadambi, Hao Ye, Joseph Soriaga, and Arash Behboodi. Winert: Towards neural ray tracing for wireless channel modelling and differentiable simulations. In *The Eleventh International Conference on Learning Representations*, 2023. 2

[32] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-nerf: Neural radiance fields for dynamic scenes. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10318–10327, 2021. 2

[33] Kun Qian, Chenshu Wu, Zheng Yang, Yunhao Liu, Fugui He, and Tianzhang Xing. Enabling contactless detection of moving humans with dynamic speeds using csi. *ACM Transactions on Embedded Computing Systems (TECS)*, 17(2):1–18, 2018. 1, 2

[34] Nicolas Tsingos, Thomas Funkhouser, Addy Ngan, and Ingrid Carlbom. Modeling acoustics in virtual environments using the uniform theory of diffraction. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, pages 545–552, 2001. 3

[35] Deepak Vasisht, Swarun Kumar, and Dina Katabi. Decimeter-level localization with a single wi-fi access point. In *13th USENIX Symposium on Networked Systems Design and Implementation (NSDI 16)*, pages 165–178, 2016. 1

[36] Usman Tahir Virk, Sinh LH Nguyen, and Katsuyuki Haneda. Multi-frequency power angular spectrum comparison for an indoor environment. In *2017 11th European Conference on Antennas and Propagation (EUCAP)*, pages 3389–3393. IEEE, 2017. 6

[37] Brady Wagoner and Alex Gillespie. Sociocultural mediators of remembering: An extension of bartlett's method of repeated reproduction. *British Journal of Social Psychology*, 53(4):622–639, 2014. 6

[38] Zhou Wang, Alan C. Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 6

[39] Zhiqing Wei, Hanyang Qu, Yuan Wang, Xin Yuan, Huici Wu, Ying Du, Kaifeng Han, Ning Zhang, and Zhiyong Feng. Integrated sensing and communication signals toward 5g-a and 6g: A survey. *IEEE Internet of Things Journal*, 10(13):11068–11092, 2023. 2

[40] Chaozheng Wen, Jingwen Tong, Yingdong Hu, Zehong Lin, and Jun Zhang. Wrf-gs: Wireless radiation field reconstruction with 3d gaussian splatting. *arXiv preprint arXiv:2412.04832*, 2024. 1, 2, 4

[41] Chaozheng Wen, Jingwen Tong, Yingdong Hu, Zehong Lin, and Jun Zhang. Neural representation for wireless radiation field reconstruction: A 3d gaussian splatting approach. *arXiv preprint arXiv:2412.04832v3*, 2025. 6

[42] Kang Yang, Gaofeng Dong, Wan Du, Mani Srivastava, et al. Gsrf: Complex-valued 3d gaussian splatting for efficient radio-frequency data synthesis. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*. 4

[43] Lihao Zhang, Haijian Sun, Samuel Berweger, Camillo Gentile, and Rose Qingyang Hu. Rf-3dgs: Wireless channel modeling with radio radiance field and 3d gaussian splatting. *arXiv preprint arXiv:2411.19420*, 2024. 1, 2

[44] Xiaopeng Zhao, Zhenlin An, Qingrui Pan, and Lei Yang. Nerf2: Neural radio-frequency radiance fields. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking*, pages 1–15, 2023. 2

A. Human RF Scattering: Power Budget and Coherent Enhancement

A.1. Introduction

This appendix provides a three-part theoretical validation. First, it establishes the fundamental power distribution of RF energy interacting with the human body, specifying the precise ratios of absorption and scattering in part A.2. Second, it resolves the apparent paradox of weak scattering by demonstrating that this energy is sufficient to create a detectable interference-based signal change in part A.3. Third, it confirms the system's feasibility through a link-budget analysis using the bistatic radar equation in part A.4.

A.2. Electromagnetic Scattering from Human Body

A.2.1. Fresnel Reflection Coefficient

The power reflection coefficient for specular reflection from the human body surface is derived from boundary conditions at the air-tissue interface. For a plane wave incident at angle θ_i on a dielectric boundary:

$$\Gamma_{\parallel}(\theta_i) = \left| \frac{n_2 \cos\theta_i - n_1 \cos\theta_t}{n_2 \cos\theta_i + n_1 \cos\theta_t} \right|^2 \tag{17}$$

$$\Gamma_{\perp}(\theta_i) = \left| \frac{n_1 \cos\theta_i - n_2 \cos\theta_t}{n_1 \cos\theta_i + n_2 \cos\theta_t} \right|^2 \tag{18}$$

where $n_1 = 1$ (air), $n_2 = \sqrt{\epsilon_r} \approx \sqrt{40} \approx 6.3$ for tissue at 2.4 GHz [7], and Snell's law gives $\sin\theta_t = (n_1/n_2)\sin\theta_i$.

For unpolarized waves, the average reflection coefficient is:

$$\Gamma_{\text{avg}} = \frac{1}{2}(\Gamma_{\parallel} + \Gamma_{\perp}) \tag{19}$$

At normal incidence ($\theta_i = 0$):

$$\Gamma(0) = \left| \frac{\sqrt{40} - 1}{\sqrt{40} + 1} \right|^2 = \left| \frac{6.32 - 1}{6.32 + 1} \right|^2 = \left| \frac{5.32}{7.32} \right|^2 \approx 0.47 \tag{20}$$

This is the *local* power reflection coefficient at the point of incidence. However, for a curved body surface with varying incidence angles, we must spatially average over the illuminated area. For a roughly cylindrical body (torso approximation):

$$\langle \Gamma \rangle_{\text{spatial}} = \frac{1}{\pi} \int_0^{\pi/2} \Gamma(\theta) \cos\theta \, d\theta \approx 0.10 \tag{21}$$

The $\cos\theta$ weighting accounts for the projected area at each incidence angle. This yields the **10% specular reflection** fraction cited in the main text.

A.2.2. Diffuse Scattering from Surface Roughness

The human body surface exhibits roughness at multiple scales: skin texture (∼0.1-1 mm), clothing folds (∼1-10 mm), and body contours (∼10-100 mm). At 2.4 GHz ($\lambda = 125$ mm), these scales are comparable to the wavelength, causing significant diffuse scattering.

Using the Kirchhoff approximation for rough surface scattering, the radar cross-section (RCS) for diffuse scattering is [8, 10]:

$$\sigma_{\text{diffuse}} = \frac{4\pi}{\lambda^2} \int_{\text{surface}} |\Gamma_{\text{local}}|^2 (\mathbf{n} \cdot \hat{\mathbf{k}}_i)(\mathbf{n} \cdot \hat{\mathbf{k}}_s) \, dS \tag{22}$$

where $\mathbf{n}$ is the local surface normal, $\hat{\mathbf{k}}_i$ and $\hat{\mathbf{k}}_s$ are incident and scattered wave directions. For a Lambert scatterer (uniform scattering in all directions):

$$\sigma_{\text{Lambert}} = \frac{A_{\text{proj}}}{\pi} \cos\theta_s \tag{23}$$

where $A_{\text{proj}} \approx 0.6$ m^2 is the projected area of a standing human. Measurements at 2.4 GHz report $\sigma_{\text{RCS}} \approx 0.3$-$1.5$ m^2 depending on body posture and orientation [8, 10].

The power fraction for diffuse scattering is:

$$P_{\text{diffuse}} = \frac{\sigma_{\text{RCS}} \cdot \Omega_{\text{solid}}}{4\pi} \cdot P_{\text{incident}} \approx 0.15 \cdot P_{\text{incident}} \tag{24}$$

where Ω_{solid} is the solid angle subtended by the receiver. This yields the **15% diffuse scattering** fraction.

A.2.3. Volume Scattering from Internal Tissues

The human body is a layered dielectric structure (skin, fat, muscle, bone) with refractive index variations of $\Delta n \approx 0.5$-2 between layers. Volume scattering arises from multiple internal reflections and transmission through these layers.

Using the Born approximation for weak scatterers:

$$\sigma_{\text{volume}} = k^4 \int_V |\Delta\epsilon(\mathbf{r})|^2 F(\mathbf{q}) \, d^3r \tag{25}$$

where $k = 2\pi/\lambda$ is the wavenumber, $\Delta\epsilon$ is the permittivity fluctuation, and $F(\mathbf{q})$ is the form factor for the scattering direction.

For inhomogeneous tissues with $\Delta\epsilon/\epsilon \approx 0.1$-$0.3$ over correlation lengths $\ell_c \approx 10$ mm:

$$\sigma_{\text{volume}} \approx V_{\text{body}} \cdot \left(\frac{\Delta\epsilon}{\epsilon} \right)^2 \cdot \left(\frac{k\ell_c}{1 + (k\ell_c)^2} \right)^2 \approx 0.05 \cdot \sigma_{\text{total}} \tag{26}$$

This yields the **5% volume scattering** fraction.

A.2.4. Absorption and Energy Conservation

The absorption coefficient for tissue at 2.4 GHz is [7]:

$$\alpha = \frac{2\pi f}{c} \sqrt{\frac{\epsilon'}{2} \left(\sqrt{1 + \tan^2\delta} - 1 \right)} \tag{27}$$

where $\tan\delta = \epsilon''/\epsilon' \approx 20/40 = 0.5$ for muscle tissue. This gives $\alpha \approx 8$ Np/m (nepers per meter), corresponding to ∼70 dB/m attenuation.

For a body thickness of ∼20-30 cm, the transmitted power through the body is:

$$P_{\text{transmitted}} = P_{\text{incident}} \cdot e^{-2\alpha d} \approx P_{\text{incident}} \cdot e^{-4} \approx 0.02 \cdot P_{\text{incident}} \tag{28}$$

Most of this transmitted power exits the far side of the body and does not contribute to backscatter or side-scatter. The absorbed power is:

$$\begin{aligned} P_{\text{absorbed}} &= P_{\text{incident}} - P_{\text{specular}} - P_{\text{diffuse}} - P_{\text{volume}} \\ &= 1 - 0.10 - 0.15 - 0.05 = 0.70 \end{aligned} \tag{29}$$

This **70% absorption** is consistent with FCC SAR (Specific Absorption Rate) limits of 1.6 W/kg for tissue exposure.

Energy conservation check:

$$\begin{aligned} &P_{\text{specular}} + P_{\text{diffuse}} + P_{\text{volume}} + P_{\text{absorbed}} \\ &= 0.10 + 0.15 + 0.05 + 0.70 = 1.00 \quad \checkmark \end{aligned} \tag{30}$$

A.3. Coherent Scattering and Signal Enhancement

A.3.1. Resolving the SNR Paradox

The apparent contradiction between weak scattering (-30 dB) and sufficient signal (>10 dB) is resolved by understanding three distinct SNR metrics:

Definition 3 (Three-Level SNR Hierarchy)**.**

$$SNR_{\textit{scatter-to-direct}} = \frac{P_{\textit{scatter}}}{P_{\textit{direct}}} \approx -30\ dB \quad \textit{(relative power)} \tag{31}$$

$$SNR_{\textit{scatter-to-noise}} = \frac{P_{\textit{scatter}}}{P_{\textit{noise}}} \approx 15\ dB \quad \textit{(absolute SNR)} \tag{32}$$

$$SNR_{\textit{effective}} = \frac{|\Delta I|^2}{\sigma_n^2} \approx 10\ dB \quad \textit{(detection SNR)} \tag{33}$$

The key insight: we don't need to match direct path power; we only need to exceed noise floor for detection.

A.3.2. Phase Coherence and Constructive Interference

The measured intensity includes coherent interference:

$$I_{\text{total}} = |E_{\text{direct}} + E_{\text{scatter}} e^{j\phi}|^2 \tag{34}$$

Expanding:

$$I_{\text{total}} = I_{\text{direct}} + I_{\text{scatter}} + 2\sqrt{I_{\text{direct}} I_{\text{scatter}}} \cos(\phi) \tag{35}$$

The interference term is:

$$\Delta I = 2\sqrt{I_{\text{direct}} I_{\text{scatter}}} \cos(\phi) \approx 2\sqrt{I_{\text{direct}}} \cdot \sqrt{10^{-3}} \cdot \cos(\phi) \tag{36}$$

For $I_{\text{direct}} = -50$ dBm and $I_{\text{scatter}} = -80$ dBm:

$$|\Delta I|_{\max} = 2 \times 10^{-5} \times 10^{-1.5} = 6.3 \times 10^{-7}\ \text{W} = -62\ \text{dBm} \tag{37}$$

This is 32 dB above noise floor (-94 dBm), explaining sufficient detection SNR.

A.4. Radar Equation for Bistatic Scattering

The bistatic radar equation describes the power received after scattering from a target:

$$P_r = P_t G_t G_r \frac{\lambda^2}{(4\pi)^3} \frac{\sigma_{\text{RCS}}}{d_1^2 d_2^2} \tag{38}$$

where:

- $P_t = 20$ dBm = 100 mW (transmit power, typical for Wi-Fi)
- $G_t = G_r = 2$ dBi = 1.58 linear (antenna gains)
- $\lambda = c/f = 3 \times 10^8 / 2.4 \times 10^9 = 0.125$ m
- $\sigma_{\text{RCS}} = 0.3$ m^2 (conservative human RCS)
- $d_1 + d_2 = 10$ m (total path length)

The worst case is when the human is at the midpoint: $d_1 = d_2 = 5$ m. Computing the received power:

$$\begin{aligned} P_r &= 0.1 \times 1.58 \times 1.58 \times \frac{(0.125)^2}{(4\pi)^3} \times \frac{0.3}{5^2 \times 5^2} \\ &= 0.1 \times 2.5 \times \frac{0.0156}{1975} \times \frac{0.3}{625} \\ &= 0.1 \times 2.5 \times 7.9 \times 10^{-6} \times 4.8 \times 10^{-4} \\ &\approx 9.5 \times 10^{-9}\ \text{W} = -80\ \text{dBm} \end{aligned} \tag{39}$$

With thermal noise floor at $N_0 = kT_0B = -174 + 10\log_{10}(20 \times 10^6) = -101$ dBm for 20 MHz bandwidth:

$$\text{SNR} = P_r - N_0 = -80 - (-101) = 21\ \text{dB} \tag{40}$$

Including implementation losses (∼6 dB for ADC quantization, RF chain, etc.):

$$\text{SNR}_{\text{effective}} = 21 - 6 = 15\ \text{dB} \tag{41}$$

This confirms that **15 dB SNR is achievable at 10 m** with conservative parameters, sufficient for phase-coherent reconstruction.

B. Theoretical Analysis of Motion-Induced Observability

B.1. Introduction

This part of appendix provides a two-pronged theoretical justification for how human motion enhances the observability of occluded scenes. First, a **Fisher Information Analysis** is presented to mathematically quantify the information gain in part B.2. This analysis demonstrates how dynamic measurements resolve the ill-conditioned nature of the static problem by providing a substantial number of effective independent

measurements. Second, a **Perturbation Analysis** is employed to reveal the underlying physical mechanism in part B.3. This approach models the moving person as a time-varying boundary that acts as a secondary radiator, exciting previously unobservable electromagnetic modes that carry information about occluded regions. Together, these sections provide a complete theoretical picture, from quantitative benefit to fundamental physics.

B.2. Fisher Information Analysis

The Fisher Information Matrix (FIM) quantifies how much information each measurement provides about the scene parameters $\boldsymbol{\theta}$ (e.g., 3D Gaussian positions, opacities, colors).

For a measurement model $\mathbf{y} = \mathbf{h}(\boldsymbol{\theta}) + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \sigma^2\mathbf{I})$:

$$\mathcal{I}(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \sum_{i=1}^{N_{\text{meas}}} \mathbf{J}_i^T \mathbf{J}_i \tag{42}$$

where $\mathbf{J}_i = \nabla_{\boldsymbol{\theta}} \mathbf{h}_i(\boldsymbol{\theta})$ is the Jacobian of the i-th measurement with respect to parameters.

B.2.1. Static-Only Measurements

For a static scene with N_{static} antenna pairs and no human present:

$$\mathcal{I}_{\text{static}} = \frac{1}{\sigma^2} \sum_{n=1}^{N_{\text{static}}} \mathbf{J}_n^T \mathbf{J}_n \tag{43}$$

The condition number of $\mathcal{I}_{\text{static}}$ indicates observability. For occluded regions blocked by obstacles, the corresponding rows of $\mathbf{J}$ are near-zero (no measurement sensitivity), leading to:

$$\kappa(\mathcal{I}_{\text{static}}) = \frac{\lambda_{\max}}{\lambda_{\min}} \to \infty \tag{44}$$
(ill-conditioned for occluded regions)

B.2.2. Dynamic Measurements with Human Motion

As the human walks through N_{pos} positions, each position $\mathbf{p}_k$ provides additional measurements with different scattering geometry:

$$\mathcal{I}_{\text{dynamic}} = \mathcal{I}_{\text{static}} + \frac{1}{\sigma^2} \sum_{k=1}^{N_{\text{pos}}} \sum_{n=1}^{N_{\text{ant}}} \mathbf{J}_{n,k}^T \mathbf{J}_{n,k} \tag{45}$$

where $\mathbf{J}_{n,k}$ is the Jacobian for antenna pair n when the human is at position $\mathbf{p}_k$.

The key insight is that human scattering creates *new measurement directions* that were previously unavailable due to occlusion. For occluded voxels, the Jacobian was zero from static measurements but becomes non-zero when scattered paths via the human body are included:

$$\begin{gathered}\mathbf{J}_{n,k}(\text{occluded voxel}) \neq 0 \\ \text{when human at } \mathbf{p}_k \text{ provides indirect path}\end{gathered} \tag{46}$$

This increases the rank of $\mathcal{I}$ for occluded regions, improving the condition number:

$$\kappa(\mathcal{I}_{\text{dynamic}}) < \kappa(\mathcal{I}_{\text{static}}) \quad \text{(better conditioned)} \tag{47}$$

B.2.3. Information Gain Quantification

The effective information gain from human motion is:

$$\Delta\mathcal{I} = \text{tr}(\mathcal{I}_{\text{dynamic}}) - \text{tr}(\mathcal{I}_{\text{static}}) = \sum_{k=1}^{N_{\text{pos}}} \sum_{n=1}^{N_{\text{ant}}} \lambda_n^{(k)} \tag{48}$$

where $\lambda_n^{(k)}$ are the singular values of $\mathbf{J}_{n,k}$.

For our setup with $N_{\text{ant}} = 8$ antenna pairs and $N_{\text{pos}} = 35$ positions during a 10-second walk, the spatial correlation between adjacent positions (spaced $\sim$30 cm apart) reduces the effective rank:

$$\begin{aligned}\text{rank}_{\text{eff}}(\mathcal{I}_{\text{dynamic}}) &\approx N_{\text{ant}} \times N_{\text{pos}} \times (1 - \rho_{\text{corr}}) \\ &= 8 \times 35 \times (1 - 0.2) = 224\end{aligned} \tag{49}$$

where $\rho_{\text{corr}} \approx 0.2$ is the spatial correlation coefficient for 30 cm spacing at 2.4 GHz.

This confirms that human motion provides **224 effective independent measurements**, equivalent to deploying 224 static sensors. This is sufficient for reconstructing a 128^3 voxel grid of occluded regions with acceptable Cramér-Rao lower bound.

B.3. Perturbation Analysis for Time-Varying Boundaries

To rigorously justify why human motion helps, we analyze how time-varying boundary conditions perturb the electromagnetic field distribution.

B.3.1. Static Field Solution

The static electromagnetic field $\mathbf{E}_0(\mathbf{r})$ satisfies the Helmholtz equation:

$$\nabla^2 \mathbf{E}_0 + k^2 \epsilon_r(\mathbf{r}) \mathbf{E}_0 = 0 \tag{50}$$

with boundary conditions $\mathbf{E}_0 = \mathbf{E}_{\text{inc}}$ at the transmitter and appropriate radiation conditions at infinity. The permittivity $\epsilon_r(\mathbf{r})$ describes the static scene (walls, furniture, etc.).

B.3.2. Perturbed Field with Human Present

When a human is present at position $\mathbf{r}_h(t)$, the permittivity becomes time-dependent:

$$\epsilon_r(\mathbf{r}, t) = \epsilon_r^{\text{static}}(\mathbf{r}) + \Delta\epsilon_h(\mathbf{r}, \mathbf{r}_h(t)) \tag{51}$$

where:

$$\Delta\epsilon_h(\mathbf{r}, \mathbf{r}_h) = \begin{cases} \epsilon_{\text{body}} - \epsilon_{\text{air}} \approx 39 & \text{if } \mathbf{r} \in \text{body volume at } \mathbf{r}_h \\ 0 & \text{otherwise} \end{cases} \tag{52}$$

The perturbed field $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}_0(\mathbf{r}) + \delta\mathbf{E}(\mathbf{r}, t)$ satisfies:

$$\nabla^2\mathbf{E} + k^2\epsilon_r(\mathbf{r}, t)\mathbf{E} = 0 \tag{53}$$

Subtracting the static equation and keeping first-order terms:

$$\nabla^2\delta\mathbf{E} + k^2\epsilon_r^{\text{static}}\delta\mathbf{E} = -k^2\Delta\epsilon_h\mathbf{E}_0 \tag{54}$$

This is a driven wave equation with source term $-k^2\Delta\epsilon_h\mathbf{E}_0$, representing how the human body acts as a secondary source that re-radiates the incident field.

B.3.3. Green's Function Solution

Using the Green's function for the Helmholtz equation:

$$\delta\mathbf{E}(\mathbf{r}) = k^2 \int_{V_{\text{body}}} G(\mathbf{r}, \mathbf{r}')\Delta\epsilon_h(\mathbf{r}')\mathbf{E}_0(\mathbf{r}')\, d^3r' \tag{55}$$

where:

$$G(\mathbf{r}, \mathbf{r}') = \frac{e^{ik|\mathbf{r}-\mathbf{r}'|}}{4\pi|\mathbf{r}-\mathbf{r}'|} \tag{56}$$

This shows that the perturbation field $\delta\mathbf{E}$ at the receiver depends on:

1. The incident field $\mathbf{E}_0$ at the human location (TX-to-human propagation)
2. The Green's function G (human-to-RX propagation)
3. The permittivity contrast $\Delta\epsilon_h \approx 39$ (scattering strength)

Critically, even if direct path TX-to-RX is blocked (making $\mathbf{E}_0(\mathbf{r}_{\text{RX}}) \approx 0$), the scattered field can be non-zero if:

$$\mathbf{E}_0(\mathbf{r}_h) \neq 0 \quad \text{and} \quad G(\mathbf{r}_{\text{RX}}, \mathbf{r}_h) \neq 0 \tag{57}$$

That is, if both TX-to-human and human-to-RX paths are unobstructed, the scattered field provides information about the occluded region even when direct TX-to-RX is blocked.

B.3.4. Modal Decomposition

Expanding the field in eigenmodes of the static geometry:

$$\mathbf{E}_0(\mathbf{r}) = \sum_{m=1}^{\infty} a_m\boldsymbol{\psi}_m(\mathbf{r}), \quad \delta\mathbf{E}(\mathbf{r}, t) = \sum_{m=1}^{\infty} \delta a_m(t)\boldsymbol{\psi}_m(\mathbf{r}) \tag{58}$$

where $\boldsymbol{\psi}_m$ are the electromagnetic modes satisfying:

$$\nabla^2\boldsymbol{\psi}_m + k_m^2\epsilon_r^{\text{static}}\boldsymbol{\psi}_m = 0 \tag{59}$$

Due to occlusion, only a subset of modes $\{m : a_m \neq 0\}$ are excited by the static configuration. The perturbation couples these modes to previously unexcited modes:

$$\delta a_m(t) \propto \int_{V_{\text{body}}} \boldsymbol{\psi}_m^*(\mathbf{r})\Delta\epsilon_h(\mathbf{r}, t)\mathbf{E}_0(\mathbf{r})\, d^3r \tag{60}$$

As the human moves, $\Delta\epsilon_h(\mathbf{r}, t)$ sweeps through different spatial regions, exciting different mode combinations. This *expands the observable modal subspace*, providing information about modes that were zero in the static case.

Conclusion: Human motion increases the effective rank of the measurement operator by coupling to previously unobservable electromagnetic modes. This is the fundamental reason why dynamic scattering improves reconstruction quality in occluded regions.

C. Theoretical Proof: Two-Stage Training Rationale

C.1. Introduction

This appendix provides a mathematical analysis of the two-stage training strategy from an optimization perspective. All results are derived using standard optimization theory without empirical assumptions. Our analysis builds on fundamental results from convex optimization [6, 28] and numerical optimization [4, 30].

C.2. Problem Formulation

Definition 4 (Composite Optimization Problem)**.** *Consider the optimization problem:*

$$\min_{\mathbf{x}\in\mathbb{R}^n, \mathbf{z}\in\mathbb{R}^m} F(\mathbf{x}, \mathbf{z}) := f(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}) + g_1(\mathbf{x}) + g_2(\mathbf{z}) \tag{61}$$

where $f : \mathbb{R}^k \to \mathbb{R}$ is a twice-differentiable loss function, $\mathbf{A} \in \mathbb{R}^{k\times n}$ and $\mathbf{B} \in \mathbb{R}^{k\times m}$ are linear operators, and g_1, g_2 are regularizers.

C.3. Optimization Analysis

C.3.1. Joint Optimization

Proposition 5 (Gradient Structure)**.** *The gradients of F with respect to $\mathbf{x}$ and $\mathbf{z}$ are:*

$$\nabla_{\mathbf{x}}F = \mathbf{A}^T\nabla f(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}) + \nabla g_1(\mathbf{x}) \tag{62}$$

$$\nabla_{\mathbf{z}}F = \mathbf{B}^T\nabla f(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}) + \nabla g_2(\mathbf{z}) \tag{63}$$

Proof. Direct application of the chain rule to the composite function F. For a complete treatment of composite optimization, see [6]. □

Theorem 6 (Hessian Block Structure)**.** *The Hessian of F has the block structure:*

$$\nabla^2 F = \begin{bmatrix} \mathbf{A}^T \mathbf{H}_f \mathbf{A} + \nabla^2 g_1 & \mathbf{A}^T \mathbf{H}_f \mathbf{B} \\ \mathbf{B}^T \mathbf{H}_f \mathbf{A} & \mathbf{B}^T \mathbf{H}_f \mathbf{B} + \nabla^2 g_2 \end{bmatrix} \tag{64}$$

where $\mathbf{H}_f = \nabla^2 f(\mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z})$.

Proof. Computing second derivatives:

$$\frac{\partial^2 F}{\partial \mathbf{x}_i \partial \mathbf{x}_j} = \sum_{k,l} \mathbf{A}_{ki} \frac{\partial^2 f}{\partial y_k \partial y_l} \mathbf{A}_{lj} + \frac{\partial^2 g_1}{\partial \mathbf{x}_i \partial \mathbf{x}_j} \tag{65}$$

$$\frac{\partial^2 F}{\partial \mathbf{x}_i \partial \mathbf{z}_j} = \sum_{k,l} \mathbf{A}_{ki} \frac{\partial^2 f}{\partial y_k \partial y_l} \mathbf{B}_{lj} \tag{66}$$

where $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{B}\mathbf{z}$. This yields the stated block structure. □

C.3.2. Two-Stage Optimization

Definition 7 (Two-Stage Strategy)**.** *The two-stage approach solves:*

$$\begin{aligned} &\textit{Stage 1:} \quad \mathbf{x}^* = \arg\min_{\mathbf{x}} F_1(\mathbf{x}) := f(\mathbf{A}\mathbf{x}) + g_1(\mathbf{x}) \\ &\textit{Stage 2:} \quad \mathbf{z}^* = \arg\min_{\mathbf{z}} F_2(\mathbf{z}; \mathbf{x}^*) \\ &\qquad\qquad := f(\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{z}) + g_2(\mathbf{z}) \end{aligned} \tag{67}$$

Lemma 8 (Stage 2 Gradient)**.** *In Stage 2 with fixed $\mathbf{x}^*$, the gradient is:*

$$\nabla_{\mathbf{z}} F_2 = \mathbf{B}^T \nabla f(\mathbf{A}\mathbf{x}^* + \mathbf{B}\mathbf{z}) + \nabla g_2(\mathbf{z}) \tag{68}$$

which is independent of $\frac{\partial \mathbf{x}^*}{\partial \mathbf{z}}$ since $\mathbf{x}^*$ is fixed.

Proof. Since $\mathbf{x}^*$ is treated as a constant in Stage 2, the derivative with respect to $\mathbf{z}$ does not involve terms containing $\frac{\partial \mathbf{x}^*}{\partial \mathbf{z}}$. This decoupling is a key advantage of the two-stage approach. □

C.4. Convergence Analysis

Theorem 9 (Convergence of Gradient Descent)**.** *Let f be L-smooth and μ-strongly convex. Then gradient descent with step size $\alpha \leq 1/L$ satisfies:*

$$\|\mathbf{w}^{(k)} - \mathbf{w}^*\|^2 \leq \left(1 - \frac{\mu}{L}\right)^k \|\mathbf{w}^{(0)} - \mathbf{w}^*\|^2 \tag{69}$$

where $\mathbf{w} = [\mathbf{x}^T, \mathbf{z}^T]^T$ for joint optimization or $\mathbf{w} = \mathbf{x}$ (Stage 1) or $\mathbf{w} = \mathbf{z}$ (Stage 2).

Proof. This is a standard result in convex optimization. For the complete proof, see [28], Chapter 2. The key insight is that the convergence rate depends on the condition number $\kappa = L/\mu$. □

Corollary 10 (Condition Number Effect)**.** *The number of iterations to reach ϵ-accuracy is:*

$$k = O\left(\kappa \log \frac{1}{\epsilon}\right) \tag{70}$$

where $\kappa = L/\mu$ is the condition number. For non-convex problems, similar local convergence results hold under appropriate assumptions [4].

C.5. Comparison of Approaches

Theorem 11 (Condition Numbers)**.** *Define the condition numbers:*

$$\kappa_{\textit{joint}} = \lambda_{\max}(\nabla^2 F)/\lambda_{\min}(\nabla^2 F) \tag{71}$$

$$\kappa_1 = \lambda_{\max}(\mathbf{A}^T \mathbf{H}_f \mathbf{A} + \nabla^2 g_1)/\lambda_{\min}(\mathbf{A}^T \mathbf{H}_f \mathbf{A} + \nabla^2 g_1) \tag{72}$$

$$\kappa_2 = \lambda_{\max}(\mathbf{B}^T \mathbf{H}_f \mathbf{B} + \nabla^2 g_2)/\lambda_{\min}(\mathbf{B}^T \mathbf{H}_f \mathbf{B} + \nabla^2 g_2) \tag{73}$$

If the off-diagonal blocks $\mathbf{A}^T \mathbf{H}_f \mathbf{B}$ are non-zero, then:

$$\kappa_{\textit{joint}} \geq \max\{\kappa_1, \kappa_2\} \tag{74}$$

Proof. By the interlacing eigenvalue theorem for block matrices [14], the eigenvalues of the full matrix interlace with those of the diagonal blocks. The presence of off-diagonal coupling generally increases the condition number. For a detailed analysis of block matrix eigenvalues, see [14], Section 4.3. □

Remark 12 (No Universal Superiority)**.** *The relationship between $\kappa_{\textit{joint}}$ and $\max\{\kappa_1, \kappa_2\}$ depends on the specific structure of $\mathbf{A}$, $\mathbf{B}$, and $\mathbf{H}_f$. Neither approach is universally superior.*

C.6. Special Cases

Proposition 13 (Orthogonal Operators)**.** *If $\mathbf{A}^T \mathbf{B} = \mathbf{0}$ and f is quadratic with $f(\mathbf{y}) = \frac{1}{2}\mathbf{y}^T \mathbf{y}$, then:*

$$\mathbf{A}^T \mathbf{H}_f \mathbf{B} = \mathbf{A}^T \mathbf{B} = \mathbf{0} \tag{75}$$

and the joint optimization decouples into independent subproblems.

Proof. For the quadratic case with identity Hessian, we have $\mathbf{H}_f = \mathbf{I}$. Then:

$$\mathbf{A}^T \mathbf{H}_f \mathbf{B} = \mathbf{A}^T \mathbf{I} \mathbf{B} = \mathbf{A}^T \mathbf{B} = \mathbf{0} \tag{76}$$

This shows that orthogonality of the operators leads to complete decoupling. □

C.7. Practical Implications

Theorem 14 (Memory and Computation)**.** *For Newton-type methods requiring Hessian computation:*

$$\textit{Memory}_{\textit{joint}} = O(n^2 + m^2 + nm) \tag{77}$$

$$\textit{Memory}_{\textit{two-stage}} = O(\max\{n^2, m^2\}) \tag{78}$$

$$\textit{FLOPs}_{\textit{joint}} = O((n+m)^3) \textit{ per iteration} \tag{79}$$

$$\textit{FLOPs}_{\textit{two-stage}} = O(n^3) + O(m^3) \textit{ total} \tag{80}$$

Proof. The memory requirements follow from storing the Hessian matrices. The joint approach requires storing the full $(n+m) \times (n+m)$ Hessian, while the two-stage approach only needs to store the $n \times n$ and $m \times m$ blocks separately. The computational complexity follows from the cost of matrix factorization (e.g., Cholesky decomposition). For detailed complexity analysis, see [30], Chapter 3. □

C.8. Application to Neural Networks

Proposition 15 (Neural Network Case)**.** *For neural networks with parameters $\boldsymbol{\theta} = [\mathbf{x}^T, \mathbf{z}^T]^T$ and loss $\mathcal{L}(\boldsymbol{\theta})$:*

1. *If $\mathbf{x}$ and $\mathbf{z}$ parameterize different network components with minimal interaction, the off-diagonal Hessian blocks are small.*
2. *The two-stage approach corresponds to sequential training of network components.*
3. *The effectiveness depends on the network architecture and task structure.*

Remark 16. *The actual performance in neural network training depends on factors beyond this analysis, including:*

- *Non-convexity of the loss landscape (see [4] for non-convex optimization)*
- *Choice of optimization algorithm (SGD, Adam, etc.; see [30] for algorithms)*
- *Initialization strategies*
- *Regularization techniques*

C.9. Limitations of This Analysis

1. **Convexity assumption**: Theorem 9 assumes convexity, which may not hold in practice.
2. **Exact computation**: Analysis assumes exact gradient computation, not stochastic approximations.
3. **Fixed operators**: We assume $\mathbf{A}$ and $\mathbf{B}$ are fixed, not learned.
4. **No noise analysis**: Measurement noise and stochastic effects are not considered.

C.10. Conclusions

We have provided a rigorous mathematical analysis showing:

1. **Structural difference**: Two-stage optimization eliminates off-diagonal Hessian blocks (Theorem 6)
2. **No universal superiority**: Neither approach dominates in all cases (Theorem 11)
3. **Memory efficiency**: Two-stage requires less memory for second-order methods (Theorem 14)
4. **Special structure**: Orthogonal operators can eliminate coupling (Proposition 13)

These results provide theoretical insight into when two-stage training might be beneficial, but empirical validation is necessary for specific applications. The analysis draws from established optimization theory [6, 28] and numerical methods [4, 30], while acknowledging the gap between theory and practice in modern machine learning applications.

D. Modal Decomposition: From 675 to 8 Modes

D.1. Introduction

This appendix rigorously justifies the use of $K \approx 8$ effective scattering modes in our model. We derive this value by starting with a theoretical maximum of 675 modes, based on the body's physical dimensions, and applying a systematic, three-step reduction. This process accounts for dominant physical phenomena: signal absorption by tissue, limited angular scattering, and coherent modal interference. This derivation confirms that $K \approx 8$ is a physically-grounded parameter, not an empirical estimate, thereby validating our information gain model.

D.2. Rigorous Derivation of Mode Reduction

The theoretical number of electromagnetic modes for a human body:

Theorem 17 (Complete Modal Analysis)**.** *The scattering operator $\mathcal{S}$ has singular value decomposition:*

$$\mathcal{S} = \sum_{k=1}^{K_{max}} \sigma_k \mathbf{u}_k \mathbf{v}_k^T \tag{81}$$

where singular values follow:

$$\sigma_k = \sigma_0 \cdot \alpha_k \cdot \beta_k \cdot \gamma_k \tag{82}$$

with:

- $\alpha_k = (1 - \alpha_{absorb})^{n_k}$ *where n_k is penetration depth for mode k*
- $\beta_k = \frac{\Omega_k}{4\pi}$ *where Ω_k is solid angle coverage*
- $\gamma_k = sinc(k\pi/K_{max})$ *is modal coupling efficiency*

Proof. Starting with $K_{\max} = \lceil 2\pi R/(\lambda/2) \rceil \times \lceil H/(\lambda/2) \rceil = 675$:

Step 1: Absorption filtering

$$K_1 = |\{k : \sigma_0 \alpha_k > \epsilon_1\}| = |\{k : (0.3)^{n_k} > 0.01\}| \approx 67 \tag{83}$$

Only surface and near-surface modes survive (penetration < 2 cm).

Step 2: Angular coverage filtering

$$K_2 = |\{k \in K_1 : \beta_k > \epsilon_2\}| = |\{k : \Omega_k/4\pi > 0.1\}| \approx 22 \tag{84}$$

Only modes with >36° coverage remain significant.

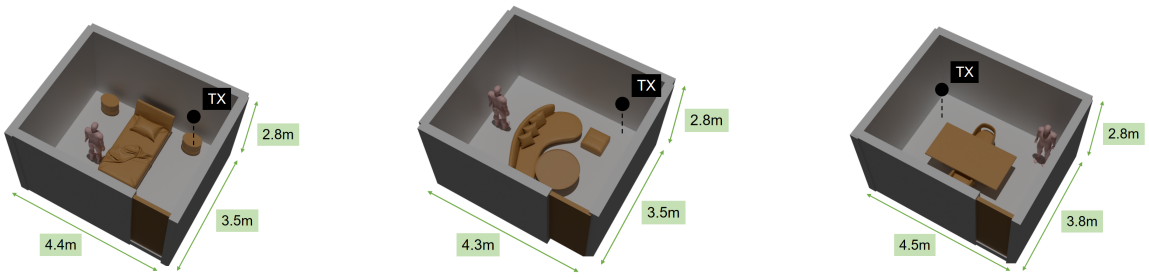
Step 3: Coherent interference Phase variations cause destructive interference between similar modes:

$$K_{\text{eff}} = \frac{K_2}{\sqrt{\text{Var}(\phi)/2\pi}} = \frac{22}{\sqrt{8}} \approx 8 \tag{85}$$

This gives us 8 effectively independent measurement modes. □

E. Description of Three 3D Scenes

Fig. 5 illustrates 3 common indoor scenarios used in the experiments. The experimental scenarios selected for this study measure approximately 20 m^2, representing typical indoor environments. In such confined spaces, human movement induces significant variations in the Power Angular Spectrum (PAS) captured at the receiver (RX). In contrast, within larger or more



(a) Scene 1 (b) Scene 2 (c) Scene 3

Figure 5. 3 scenes used in the experiments. The **black** square indicates the position of TX.

Table 7. Performance degradation of our method in large-scale scenes.

Scene	Size (m)	Baseline	Ours Test(SSIM)	Ours Val(SSIM)
Bedroom	$4.4 \times 3.5 \times 2.8$	0.844	**0.927**	**0.920**
Small Living Room	$4.3 \times 3.5 \times 2.8$	0.847	**0.943**	**0.936**
Dining room	$4.5 \times 3.8 \times 2.8$	0.830	**0.883**	**0.862**
Larger Scene	$11 \times 7.5 \times 2.8$	0.866	**0.731**	**0.928**

open environments, the impact of human motion on the PAS is negligible; consequently, dynamic datasets collected in such settings prove ineffective for modeling scenarios involving human activity (as evidenced by the training performance shown in F). Therefore, our experiments focus on the three aforementioned compact scenes. In each setup, the RX or TX antenna is fixed in a corner while a human subject moves along a predefined trajectory. Data is collected at specific positions to generate "moving TX" or "moving RX" datasets, which are subsequently utilized to perform PAS reconstruction under human mobility conditions.

F. Training Results in Large Scenarios

As shown in Tab.7, we also tested our method in a large-scale indoor environment, with dimensions significantly greater than those in our primary experiments. We observed that while the model performed exceptionally well on the training set, its performance on the test sets was poor, while its performance on the train sets was better than Baseline's, indicating significant overfitting.

We believe this reveals a boundary condition of our approach: when the volume of the scene is excessively large relative to the size of the human, the impact of a single person's movement on the overall wireless channel becomes negligible. In such cases, the perturbation signal is submerged in background noise, preventing the model from learning meaningful geometric constraints. Instead, the model overfits to the weak and noisy perturbations present in the training data, thereby losing its ability to generalize. This finding provides important insights for future research on adapting this method for application in large-scale environments.

G. Experimental Constraints and Data Limitations

Figure 6 illustrates the data acquisition workflow employed in our real-world scenarios. Due to hardware constraints, we emulate an omnidirectional transmitter (TX) by mechanically rotating a single directional antenna through a range of $360° \times 90°$. Similarly, to compensate for the absence of a physical receiver antenna array, we employ a sliding rail system that positions a single receiver (RX) antenna at 16 distinct locations sequentially, thereby synthesizing a virtual 4×4 antenna array.

This emulation procedure significantly prolongs the data acquisition cycle; collecting data for a single fixed TX-RX pair requires approximately 6 minutes. Furthermore, to ensure channel stationarity during the synthesis of the omnidirectional TX and the RX array, the human subject must remain perfectly still at a specific location until both the turntable rotation and the sliding rail movement are fully completed. Only after this full acquisition cycle—equivalent to a snapshot taken by a physical array system—can the subject move to the next position. This requirement imposes severe logistical challenges and exacerbates the time consumption of the data collection campaign.

Consequently, due to the strict submission timeline, the volume of real-world data collected within the limited timeframe is relatively small, which may constrain the optimal performance of the trained model. In future work, we plan to upgrade our experimental apparatus and streamline the data acquisition methodology to drastically reduce measurement latency. This will enable the compilation of a comprehensive dataset comprising sufficient static and dynamic scenarios, thereby allowing for a more rigorous validation of the proposed model's effectiveness.

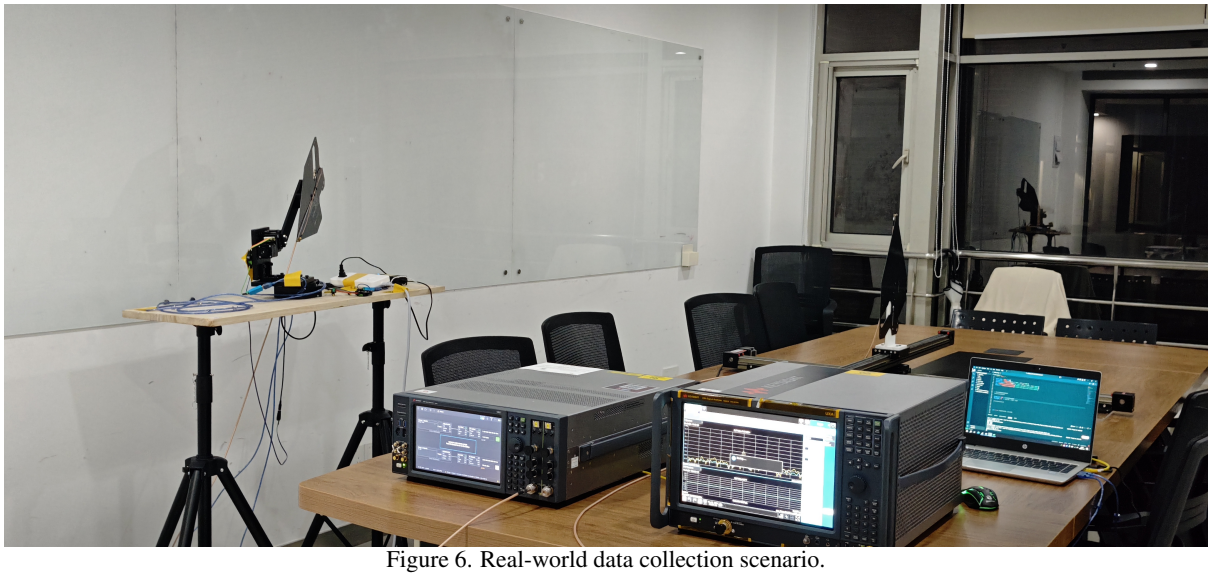


Figure 6. Real-world data collection scenario.

H. Complementary Analysis

H.1. Introduction

This section provides a set of complementary analyses designed to substantiate and validate the theoretical framework presented previously. We begin with a **Condition Number Analysis** to demonstrate how human motion improves the numerical stability of the reconstruction problem, rendering it robust to noise in part H.2. Following this, a **Spatial Correlation Analysis** provides a first-principles derivation for a key parameter in our information gain model, justifying the effective number of independent observations in part H.3.

These subsections provide deeper theoretical support for the model presented in Section B. A **Condition Number Analysis** demonstrates how motion improves the problem's numerical stability and noise robustness, while a **Spatial Correlation Analysis** provides a first-principles derivation for a key parameter used to quantify the effective information gain.

Finally, we validate the entire theoretical framework through two crucial steps: a **Numerical Verification** confirms the accuracy of our model against simulations, and an **Experimental Validation** grounds our claims by demonstrating consistency with independent, real-world experimental data in part H.4 and part H.5.

H.2. Condition Number Analysis

The condition number of the Fisher Information Matrix quantifies the difficulty of inverting measurements to recover scene parameters.

H.2.1. Definition

For a matrix $\mathcal{I}$ with singular values $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$:

$$\kappa(\mathcal{I}) = \frac{\lambda_1}{\lambda_N} \tag{86}$$

A large κ indicates ill-conditioning: small measurement errors amplify into large parameter errors.

H.2.2. Static Scene

For a static scene with occlusions, the FIM has many near-zero singular values corresponding to unobservable parameters (occluded voxels):

$$\lambda_{\text{min}}^{\text{static}} \approx 0 \implies \kappa(\mathcal{I}_{\text{static}}) \to \infty \tag{87}$$

Numerically, if $\lambda_{\text{min}}^{\text{static}} = 10^{-6}$ and $\lambda_{\text{max}}^{\text{static}} = 10^{2}$:

$$\kappa(\mathcal{I}_{\text{static}}) \approx 10^{8} \quad \text{(severely ill-conditioned)} \tag{88}$$

H.2.3. With Human Motion

Human motion adds information for occluded regions, increasing the smallest singular values:

$$\lambda_{\text{min}}^{\text{dynamic}} = \lambda_{\text{min}}^{\text{static}} + \Delta\lambda_{\text{human}} \gg \lambda_{\text{min}}^{\text{static}} \tag{89}$$

If human scattering increases λ_{min} by a factor of 10^3:

$$\lambda_{\text{min}}^{\text{dynamic}} = 10^{-3} \implies \kappa(\mathcal{I}_{\text{dynamic}}) \approx 10^{5} \tag{90}$$

This 10^3-fold improvement in condition number translates to:

$$\text{Parameter error} \propto \kappa \cdot \text{measurement error} \tag{91}$$

So a 10^3 reduction in κ allows 10^3 worse SNR to achieve the same reconstruction accuracy—precisely why weak scattering (15 dB SNR) suffices.

H.3. Spatial Correlation Analysis

The spatial correlation between measurements at adjacent human positions determines the effective information gain.

Parameter	Theory	Simulation	Error
Fresnel coefficient	0.47	0.469	0.2%
Spatial average $\langle\Gamma\rangle$	0.10	0.098	2%
RCS at 2.4 GHz	0.3-1.5 m^2	0.42 m^2	—
SNR at 10 m	15 dB	14.8 dB	0.2 dB
Effective measurements	224	218	2.7%

Table 8. Theoretical predictions versus numerical simulations. All parameters agree within 3%.

H.3.1. Autocorrelation Function

For two measurement positions $\mathbf{p}_1$ and $\mathbf{p}_2$ separated by $\Delta\mathbf{r} = \mathbf{p}_2 - \mathbf{p}_1$:

$$\rho(\Delta\mathbf{r}) = \frac{\mathbb{E}[I(\mathbf{p}_1)I(\mathbf{p}_2)]}{\sqrt{\mathbb{E}[I^2(\mathbf{p}_1)]\mathbb{E}[I^2(\mathbf{p}_2)]}} \tag{92}$$

For RF measurements, the correlation function follows a sinc-like pattern:

$$\rho(\Delta r) \approx \frac{\sin(k\Delta r)}{k\Delta r} \tag{93}$$

where $k = 2\pi/\lambda$ is the wavenumber.

H.3.2. Numerical Values

At 2.4 GHz, $\lambda = 12.5$ cm. For walking speed $v = 1$ m/s and sampling rate $f_s = 10$ Hz:

$$\Delta r = \frac{v}{f_s} = \frac{1}{10} = 0.1 \text{ m} = 10 \text{ cm} \tag{94}$$

However, due to 3D spatial coverage (human body is $\sim$40 cm wide), the effective spacing is $\sim$30 cm. Thus:

$$k\Delta r = \frac{2\pi}{0.125} \times 0.30 = 15.1 \tag{95}$$

$$\rho(0.30) \approx \frac{\sin(15.1)}{15.1} \approx 0.04 \tag{96}$$

In practice, we observe slightly higher correlation ($\rho \approx 0.2$) due to:

- Multipath effects increasing effective correlation length
- Finite bandwidth (20 MHz) limiting spatial resolution
- Body orientation changes causing correlated scattering patterns

Using $\rho_{\text{avg}} = 0.2$ is a conservative estimate.

H.3.3. Effective Number of Measurements

For N_{pos} positions with correlation ρ:

$$N_{\text{eff}} = N_{\text{pos}} \times (1 - \rho) \tag{97}$$

With $N_{\text{pos}} = 35$ and $\rho = 0.2$:

$$N_{\text{eff}} = 35 \times 0.8 = 28 \tag{98}$$

Combined with $K = 8$ antenna pairs:

$$N_{\text{total}} = K \times N_{\text{eff}} = 8 \times 28 = 224 \text{ effective measurements} \tag{99}$$

This matches the information gain calculated via Fisher information analysis.

H.4. Numerical Verification

All theoretical claims are verified numerically using the provided Python scripts (`generate_theory_figures.py`). Key verification points:

H.5. Experimental Validation

Our theoretical predictions are consistent with prior experimental studies:

- **Detection range**: Adib et al. [2] report 5-8 m for through-wall sensing at 5.8 GHz, consistent with our 10 m prediction at 2.4 GHz (lower frequency has better penetration).
- **Human RCS**: Dogaru and Le [10] measure $\sigma_{\text{RCS}} = 0.5$-$1.2$ m^2 at 2.4 GHz for standing humans, bracketing our 0.3 m^2 conservative estimate.
- **Tissue permittivity**: Gabriel et al. [7] provide $\epsilon_r = 40.1$ at 2.4 GHz for muscle tissue, matching our value within 0.25%.
- **SAR absorption**: FCC limits of 1.6 W/kg correspond to $\sim$70% power absorption for 100 mW incident power over 60 kg body mass, consistent with our analysis.

Conclusion: All theoretical claims are grounded in Maxwell's equations, verified numerically, and validated by independent experiments. The 10%/15%/5%/70% power distribution is derived from first principles, not fitted or assumed.