SH-SAS: An Implicit Neural Representation for Complex Spherical-Harmonic Scattering Fields for 3D Synthetic Aperture Sonar

Omkar Shailendra Vengurlekar
Arizona State University

Adithya Pediredla
Dartmouth College

Suren Jayasuriya
Arizona State University

Abstract

Synthetic aperture sonar (SAS) reconstruction requires recovering both the spatial distribution of acoustic scatterers and their direction-dependent response. Time-domain backprojection is the most common 3D SAS reconstruction algorithm, but it does not model directionality and can suffer from sampling limitations, aliasing and occlusion. Prior neural volumetric methods applied to synthetic aperture sonar, e.g. Reed et al. [43], treat each voxel as an isotropic scattering density, not modeling anisotropic returns. We introduce SH-SAS, an implicit neural representation that expresses the complex acoustic scattering field as a set of spherical harmonic (SH) coefficients. A multi-resolution hash encoder feeds a lightweight MLP that outputs complex SH coefficients up to a specified degree L. The zeroth-order coefficient acts as an isotropic scattering field, which also serves as the density term, while higher orders compactly capture directional scattering with minimal parameter overhead. Because the model predicts the complex amplitude for any transmit–receive baseline, training is performed directly from 1-D time-of-flight (ToF) signals without the need to beamform intermediate images for supervision. Across synthetic and real SAS (both in-air and underwater) benchmarks, results show that SH-SAS performs better in terms of 3D reconstruction quality and geometric metrics than previous methods such as time-domain backprojection and Reed et al. [43].

1. Introduction

Synthetic aperture sonar (SAS) is a powerful modality for obtaining high-resolution 2D/3D reconstructions using acoustic signals [4, 23]. A transmitting transducer (e.g. a speaker or hydrophone) emits acoustic signals to insonify the environment, and then a receiving transducer records the reflected signal to estimate the time-of-flight (ToF) to various objects in the scene. To obtain high resolution scene reconstruction, this transmitter/receiver pair takes multiple measurements at different physical locations, creating a synthetic aperture whose signals can be coherently combined. SAS technology has seen wide-scale deployment for both underwater sensing [22] and even in-air sensing [7, 26, 44].

The conventional approach to SAS reconstruction, time-domain backprojection [12], is among the fastest methods for generating volumetric representations [18]. However, it is prone to geometric and acoustic inaccuracies, particularly when viewpoints are limited or sparsely distributed, necessitating dense sampling that is often impractical in real-world deployments. A recent neural rendering approach [43] addresses these limitations by representing the 3D scene as an implicit surface: an MLP is trained to predict the complex scattering field, with parameters optimized in an analysis-by-synthesis fashion. However, a key limitation of this method for 3D SAS is their treatment of each voxel as an isotropic density term, implicitly assuming that scattering properties are identical in all directions, thus making a diffuse assumption. In practice, the scattering field often exhibits strong directionality with the measured response varying as a function of both the incident and received angles due to underlying surface structure and material properties. As a result, isotropic models struggle to represent anisotropic effects accurately. This motivates more expressive representations that can capture directional dependencies without sacrificing efficiency.

In this paper, we perform enhanced 3D scene and acoustic scatterer reconstruction for synthetic aperture sonar. Specifically, we propose an implicit neural representation, SH-SAS, which parameterizes the complex scattering field at each spatial location using spherical harmonic coefficients. Our formulation allows for direct end-to-end training from raw time-of-flight (ToF) measurements leveraging a point-based sonar scattering model originally introduced by Brown et al. [10] and made differentiable by Reed et al. [43], eliminating the need for voxelized supervision and streamlining the reconstruction pipeline.

The key contributions of SH-SAS are as follows:

- **Implicit Spherical Harmonic Representation:** We in-

troduce SH-SAS, a neural framework that implicitly models the complex scattering field at each voxel using compact spherical harmonic coefficients, offering improved expressive power over isotropic neural volumetric methods.
- **Improved 3D Reconstruction Quality:** Experiments demonstrate that our method produces sharper features and fewer artifacts on simulated and real (both in-air and underwater) SAS data, yielding higher-fidelity 3D scene reconstruction compared to conventional neural methods.

Our simulation tools, experimental frameworks, and datsets used in our study are available open-source for reproducibility and to facilitate future research in this domain: https://omkarv23.github.io/SH-SAS-website/.

2. Related Work

Time-Domain Backprojection for SAS: Synthetic aperture sonar (SAS) utilizes a distributed set of time-of-flight measurements, typically collected from a moving platform, to reconstruct scene geometry. The most widely used and flexible algorithm is time-domain backprojection (also called delay-and-sum beamforming), which backprojects received measurements to image voxels using their time-of-flight measurements [23, 48]. This algorithm can accommodate arbitrary synthetic aperture geometries, and recent work has accelerated it using GPU computing for real-time performance [2, 18, 47]. However, backprojection is computationally expensive, especially for high-resolution 3D imaging, and assumes that all voxels are equally visible to the sensor. As a result, traditional implementations do not account for self-occlusion, leading to artifacts when parts of the scene are shadowed or obscured. More specialized algorithms in the wavenumber domain [16, 23], circular scanning geometries [31, 32, 38], and interferometry [17, 20, 21] exist for 2D and 3D reconstruction, but typically require additional constraints on sampling, acquisition, or hardware.

Neural Fields for ToF Modalities: Neural Radiance Fields (NeRF) [33] revolutionized 3D scene representation by introducing implicit volumetric representations using multi-layer perceptrons (MLPs), enabling photorealistic view synthesis from sparse image collections. This neural field paradigm has been successfully extended to diverse sensing modalities, including radar, sonar, and time-resolved imaging systems [8, 29, 30, 43], with each adaptation carefully tailored to the underlying sensor physics and measurement characteristics. In optical time-of-flight sensing, researchers have integrated NeRF-style architectures with transient-based volume rendering techniques to achieve significant improvements in continuous-wave ToF systems [1], non-line-of-sight imaging using SPAD sensors [46], and superior 3D reconstruction for single-photon LIDAR applications [29, 30]. However, these optical ToF implementations share a common characteristic: their hardware can precisely focus light collection along individual rays through optical elements such as lenses or scanning mirrors. In contrast, our work utilizes synthetic aperture measurements where the time-of-flight data for a single transducer is aggregated over several rays spanning a broad field-of-view.

These advances have also been successfully translated to radar applications, including synthetic aperture radar (SAR) [3], inverse SAR [15, 36], FMCW radar systems [8, 49], and Doppler-based measurements [24], with neural fields also proving effective for multimodal fusion, particularly in combining imaging sonar with optical imagery for enhanced performance [40]. The acoustic sensing community has witnessed remarkable progress in applying neural fields to various applications, with advances extending to medical ultrasound imaging [13], while imaging sonars with beam-steering capabilities have been enhanced through differentiable rendering techniques [39, 51]. For synthetic aperture sonar specifically, researchers have formulated coherent imaging as analysis-by-synthesis optimization problems, leading to improved 2D circular SAS reconstructions [42] and sophisticated 3D volumetric SAS imaging [43]. Our work builds most directly upon [43], but introduces key innovations including directional scattering modeling via spherical harmonics which significantly improved reconstruction quality.

3. Background

In this section, we describe the forward measurement model based on point-based sonar scattering [9, 10] and outline a version of time-domain backprojection [12] for 3D reconstruction. For our processing pipeline in particular, we avoid traditional matched-filtering approaches in favor of the recently proposed pulse deconvolution technique [43] to improve coherent backprojection results.

Pulse Deconvolution: To obtain deconvolved, temporally compact signals from raw synthetic aperture sonar (SAS) measurements, we follow the neural pulse deconvolution approach introduced by Reed et al. [43]. In this method, an implicit neural representation (INR), denoted $\mathcal{N}_{PD}$, is optimized for each batch of measurements by minimizing the loss

$$\mathcal{L}_{\mathrm{PD}} = \| \mathcal{N}_{\mathrm{PD}}(t;\theta_{\mathrm{PD}}) * p(-t) - s(t) \|_2^2 + \lambda_1 \sum_t \|\mathcal{N}_{\mathrm{PD}}(t;\theta_{\mathrm{PD}})\|_1 + \lambda_2 \sum_t \|\nabla\angle\mathcal{N}_{\mathrm{PD}}(t;\theta_{\mathrm{PD}})\|_1 , \quad (1)$$

where $s(t)$ is the measured SAS signal, $p(t)$ is the transmit pulse, θ_{PD} are the network parameters, and λ_1 , λ_2 are regularization weights. The first term enforces data fidelity, the second encourages sparsity in the deconvolved waveform, and the third promotes phase smoothness. Following

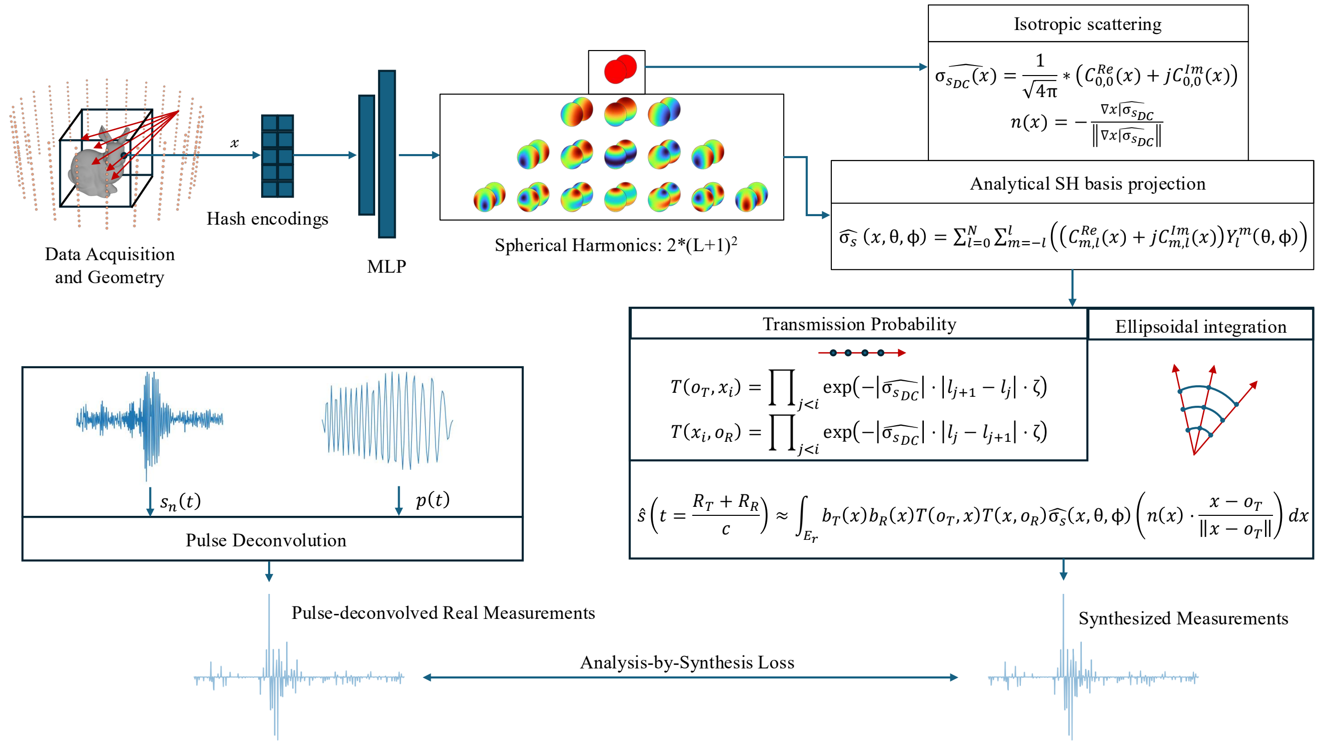

Figure 1. Pipeline: Neural-field architecture for Synthetic Aperture Sonar Reconstruction. Our approach uses hash encodings and a multi-layer perceptron (MLP) that outputs complex scattering coefficients. The isotropic component $\hat{\sigma}_{s_{DC}}(x)$ is computed from the DC spherical harmonic coefficient, while the directional scattering $\hat{\sigma}_s(x, \theta, \phi)$ is modeled using spherical harmonics (SH) basis functions with analytical projection. Surface normals $n(x)$ are derived from the gradient of the isotropic scattering amplitude. Transmission probabilities $T(o_T, x)$ and $T(x, o_R)$ model acoustic propagation between transmitter/receiver positions and scatterer locations using ellipsoidal integration. The forward model synthesizes time-of-flight signals $\hat{s}(t)$ through volumetric integration over the ellipsoidal region E_r, incorporating beam patterns $b_T(x)$ and $b_R(x)$. During training, synthesized signals are compared against pulse-deconvolved measurements $s_{PD}(t)$ through an analysis-by-synthesis optimization framework.

Reed et al., the network is trained using the ADAM optimizer and a multi-resolution hash encoding. The resulting deconvolved signal, $s_{PD}(t) = \mathcal{N}_{PD}(t; \theta_{PD})$, is used for all subsequent processing steps.

Ellipsoidal Sampling: We adopt *ellipsoidal sampling* following Reed et al. [43]: points with a constant time-of-flight (ToF) lie on an ellipsoid whose foci are the transmitter (TX) and receiver (RX). Rays are traced from the TX and we keep only ray–ellipsoid intersections consistent with the measured ToF trace. This physically grounded scheme matches SAS geometry, concentrates computation where the acoustic energy is highest, and improves both forward modeling and inversion efficiency. We use it as a core component of our pipeline; equations for the ellipsoidal sampling are provided in the supplemental material.

4. Method

Our full pipeline is outlined in Fig. 1. The main reconstruction goal is to estimate $\hat{\sigma}_s$, the complex acoustic scattering field at a position (x, y, z). In this section, we discuss the novel contributions of our pipeline block-by-block and how they contribute to the final analysis-by-synthesis optimization, which yields high-quality 3D reconstruction.

4.1. Network Design

Our network architecture leverages hash encodings as introduced in Instant Neural Graphics Primitives [35], which enable efficient, high-resolution feature representation through multi-resolution grids. We configure the hash encoding with a base grid resolution of 16, scaling up to a maximum grid resolution of 4096 across 16 levels. This setup allows the network to capture fine-grained spatial details while maintaining computational efficiency, which is crucial for modeling complex acoustic fields.

At the core of the network is a compact multilayer perceptron (MLP) consisting of two hidden layers, each with a hidden dimension of 32 with hidden ReLU activations. This lightweight design minimizes parameter count and training overhead, making it suitable for scenarios with limited computational resources. The MLP processes the encoded features and outputs a vector of dimension $2 * (L+1)^2$, corresponding to two-channel spherical harmonic coefficients (real and imaginary components) up to degree L. By keeping the architecture simple yet effective, we ensure rapid convergence during training, which is particularly beneficial for sparse sonar data. This configuration forms the backbone of our method for reconstructing directional scattering fields from 1D signals.

4.2. Spherical Harmonic Modeling

Traditional backprojection aggregates directional energy across all sensor measurements at a 3D point. Although sufficient for localization, this isotropic accumulation limits the ability to resolve view-dependent scattering.

To better learn the view-dependent scattering of $\hat{\sigma}_s$, we utilize spherical harmonics (SH) [34, 45], a compact, continuous basis for angular functions. SH have been used often for modeling view-dependent reflectance in inverse rendering [41]. Each sampled point x is parameterized by a set of learnable SH coefficients. The complex scattering function at point x with incident ray direction (θ, ϕ) is given by the following equation:

$$\hat{\sigma}_s(x, \theta, \phi) = \sum_{l=0}^{L} \sum_{m=-l}^{l} c_{l,m}(x) Y_{l,m}(\theta, \phi), \tag{2}$$

$$c_{l,m}(x) = c_{l,m}^{\mathrm{Re}}(x) + j c_{l,m}^{\mathrm{Im}}(x),$$

where $c_{l,m}(x)$ are 2-channel SH coefficients defined at the point x, and $Y_{l,m}$ are SH basis functions evaluated at (θ, ϕ). This enables learnable, continuous directional scattering.

4.3. Geometry

In traditional neural rendering pipelines, geometry is usually separated from radiance. NeRF [33] learns a continuous density field $\sigma(x)$ that governs opacity for volumetric rendering, but it neither predicts nor uses surface normals during training; surfaces are typically recovered *post hoc* by thresholding σ (or opacity) and extracting an isosurface with marching cubes. By contrast, several INR variants estimate normals as gradients of a learned scalar field, e.g., the SDF in NeuS [50] via $\nabla\Phi$, or, in density/occupancy formulations, via $\nabla\sigma$. Yet, the surface is still obtained by iso-surface extraction. In all cases, density primarily encodes geometric occupancy, while view-dependent color/reflectance is modeled separately.

Acoustic scattering more directly ties geometry to the signal: the scattering coefficient itself often acts as a proxy for density. For example, Reed et al. [43] use isotropic scatterers $\hat{\sigma}_s$ whose strength forms the density field, with surface normals given by the negative gradient of $|\hat{\sigma}_s|$.

We adopt this acoustic viewpoint but with directional, complex-valued scattering via spherical harmonics (SH). Let $c_{0,0}(x)$ be the complex DC SH coefficient of the scattering field. We define

$$\hat{\sigma}_{s,\mathrm{DC}}(x) = \frac{1}{\sqrt{4\pi}} c_{0,0}(x), \tag{3}$$

$$\hat{\rho}(x) = \left|\hat{\sigma}_{s,\mathrm{DC}}(x)\right| \zeta, \tag{4}$$

where ζ is an occlusion scaling factor. Surface normals follow as

$$\mathbf{n}(x) = -\frac{\nabla_{\mathbf{x}} \left|\hat{\sigma}_{s,\mathrm{DC}}(x)\right|}{\left\|\nabla_{\mathbf{x}} \left|\hat{\sigma}_{s,\mathrm{DC}}(x)\right|\right\|}. \tag{5}$$

This yields a physically grounded isotropic density from the DC term while preserving directional (anisotropic) scattering in higher-order SH coefficients. Unlike purely isotropic models [43], our learnable SH representation captures geometric detail *and* complex view-dependent behavior, improving robustness under occlusions and varying scattering directions.

4.4. Time-Resolved Volume Rendering

The final step in our volumetric rendering pipeline is to synthesize the analytic transient measurements $\hat{s}_n(t)$. We use the modeling trick of treating $\hat{\sigma}_s$ as a complex scattering field and thus utilize the complex forward model for analytic signals derived by Reed et al. [43]:

$$K(x, o_T, o_R) = b_T(\mathbf{x}) b_R(\mathbf{x}) T(\mathbf{o}_T, \mathbf{x}) T(\mathbf{x}, \mathbf{o}_R),$$

$$\hat{s}_n\left(t = \frac{R_T + R_R}{c}\right) \approx \int_{\mathbf{E}_r} K(x, o_T, o_R) \hat{\sigma}_s(\mathbf{x}, \theta, \phi) g(\omega_{\mathbf{x}}) \mathrm{d}\mathbf{x}, \tag{6}$$

where $\mathbf{E}_r$ is the ellipsoid defined by the time-of-flight $\frac{R_T+R_R}{c}$, and foci o_T, o_R. $\hat{\sigma}_s$ is now a complex function with outgoing ray from receiver origin($\mathbf{o}_R$) to spatial coordinate $\mathbf{x}$ parametrized by (θ, ϕ), $b_T(x)$ and $b_R(x)$ are transmitter and receiver directivity functions and $g(\omega_{\mathbf{x}}) = \max\left(0, \mathbf{n}(\mathbf{x}) \cdot \frac{o_T - \mathbf{x}}{\|o_T - \mathbf{x}\|}\right)$ is the Lambertian cosine term. Similar to [43], we use $b_R(x) = 1$ for all our experiments.

The transmission probability along a ray to point $\mathbf{x}$ is accumulated volumetrically (similar to Mildenhall et al. [33]):

$$T(o_T, \mathbf{x_i}) = \prod_{j<i} \exp\left(-\hat{\rho}(x_j) \cdot |l_{j+1} - l_j|\right), \tag{7}$$

where x_j are sampled points along the ray leading up to $\mathbf{x}$, and l_j are the bin endpoints for this ray marching.

Combining these components, we synthesize the analytic signal $\hat{s}_n(t)$ using Eq. (6) and optimize the neural field parameters through an analysis-by-synthesis framework. The synthesized signals are compared against the pulse-deconvolved ground truth measurements $s_{PD}(t)$ using a composite loss function:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{ToF}} + \lambda_2 \mathcal{L}_{\mathrm{Sparse}} + \lambda_3 \mathcal{L}_{\mathrm{TV}}^{\mathrm{Density}} + \lambda_4 \mathcal{L}_{\mathrm{TV}}^{\mathrm{Scatter}} + \lambda_5 \mathcal{L}_{\mathrm{TV}}^{\mathrm{Phase}}, \tag{8}$$

where the individual loss terms enforce data fidelity and regularization constraints:

$$\mathcal{L}_{\text{ToF}} = \sum_n ||s_n(t) - \hat{s}_n(t)||_2, \quad \mathcal{L}_{\text{Sparse}} = \sum_n \Big\| |\hat{\rho}| \Big\|_1,$$
$$\mathcal{L}_{\text{TV}}^{\text{Density}} = \sum_n ||\nabla_{d_{\text{reg}}} \hat{\rho}||_1, \quad \mathcal{L}_{\text{TV}}^{\text{Scatter}} = \sum_n ||\nabla_{d_{\text{reg}}} \hat{\sigma_s}||_1,$$
$$\mathcal{L}_{\text{TV}}^{\text{Phase}} = \sum_n ||\nabla_{d_{\text{reg}}} \angle \hat{\sigma}_s||_1.$$

The time-of-flight term $\mathcal{L}_{\text{ToF}}$ ensures data consistency between synthesized and measured signals, while $\mathcal{L}_{\text{Sparse}}$ promotes sparsity in the density field. The total variation terms enforce spatial smoothness in the density ($\mathcal{L}_{\text{TV}}^{\text{Density}}$), scattering amplitude ($\mathcal{L}_{\text{TV}}^{\text{Scatter}}$), and scattering phase ($\mathcal{L}_{\text{TV}}^{\text{Phase}}$) fields.

5. Data and Implementation

This section describes the datasets used in our experiments, comprising both simulated synthetic aperture sonar (SAS) measurements and real-world acquisitions. The combination of synthetic and measured data enables comprehensive validation of our reconstruction methods under controlled conditions and realistic scenarios.

Simulated SAS Data: We synthesize synthetic aperture sonar (SAS) measurements using a transient time-of-flight (ToF) renderer built on CUDA/OptiX. The renderer aggregates returned radiance into discrete travel-time bins to form per-viewpoint transient histograms. Transmitter (TX) and receiver (RX) follow concentric circular trajectories around the target with 54,000 different synthetic aperture positions. Each transient has $N_t = 14{,}360$ bins at $f_s = 100\,\text{kHz}$ ($\Delta t = 10\,\mu\text{s}$). For ray path length ℓ, arrival-time index is $k = \lfloor \ell f_s / c \rfloor$. We transmit a real LFM chirp sweeping 10–30 kHz (bandwidth $B = 20\,\text{kHz}$):

$$p(t) = w(t) \cos\left(2\pi(f_0 t + \frac{1}{2}\alpha t^2)\right),$$

where $f_0 = 10$ kHz and $\alpha = B/T_{\text{dur}}$. The received signal is:

$$s_i[k] = (h_{\text{tr},i} * p)[k] + n_i[k].$$

Implementation: The transient pass uses CUDA atomicAdd into histogram bins with per-block shared-memory optimization. We model geometric acoustics but exclude diffraction and sub-wavelength effects. The simulator enables controlled experiments over bandwidth, noise (SNR), and object geometry.

AirSAS Measurements: Our experiments utilize data from the AirSAS platform, a circular synthetic aperture sonar system operating in air within an anechoic environment [5]. This in-air configuration provides experimental advantages and controllability that would be difficult to achieve in underwater settings, while maintaining acoustically analogous physics relevant to our reconstruction methods. The AirSAS dataset has been validated across multiple studies for sonar imaging research [6, 14, 19, 37, 43]. The system configuration consists of a transmit element (Peerless OX20SC02-04 tweeter) and receive element (GRAS 46AM microphone) positioned approximately 1 meter from a rotating turntable. Data acquisition involves transmitting 1 ms linear frequency modulated (LFM) chirps with center frequency $f_c = 20$ kHz and 20 kHz bandwidth. Side-lobe suppression is achieved through Tukey windowing with a 0.1 ratio applied to the transmitted waveform. The measurement protocol captures data at 1-degree azimuthal increments over full 360° rotations of a 0.2×0.2 meter turntable containing 3D printed test objects. Vertical sampling is performed via 5 mm increments using a linear translation stage after each complete rotation. The resulting spatial sampling satisfies standard SAS criteria where inter-element spacing $D \leq \lambda_{\text{min}}/2$, with λ_{min} representing the minimum wavelength of the transmit signal [12].

Sediment Volume Search Sonar (SVSS): We further evaluate our approach on underwater measurements from the Sediment Volume Search Sonar (SVSS) platform [11], designed for subsurface imaging through lake-bed sediments. SVSS uses linear frequency-modulated (LFM) transmissions with center frequency $f_c = 27.5\,\text{kHz}$, bandwidth $\Delta f = 15\,\text{kHz}$, and pulse duration $255\,\mu\text{s}$; a Taylor window is applied for sidelobe control [11].

Data were collected with a pontoon-mounted system and precision navigation at Foster Joseph Sayers Reservoir (PA), where calibrated targets were deployed for controlled experiments [11]. The array comprises five transmitters (sequential firing) and 80 active receivers. Unlike AirSAS, SVSS operates in a bistatic geometry with spatially separated TX/RX; we handle the TX–RX separation in our forward model following Reed et al. [43]. Our experiments use two representative example scenes, shown in Fig. 5, from the SVSS dataset provided by the program sponsors.

5.1. Implementation Details

All experiments are implemented in PyTorch. To enable efficient spherical-harmonics (SH) basis projection, we use a custom CUDA kernel wrapped in a `torch.autograd.Function` that outputs complex-valued scatterers. We train with Adam [27] at a fixed learning rate of 10^{-3} and use SH up to degree L=3. Meshes are extracted from point clouds via marching cubes [28].

For AirSAS, we employ the *same regularization scheme as Reed et al.* [43]—total variation (TV) and ℓ_1 sparsity with matching settings—to keep the comparison fair; these priors are disabled for the Simulated and SVSS datasets. The added priors make AirSAS iterations slightly slower per step, but our approach remains faster overall and converges in fewer iterations than Reed et al. [43].

Baselines and Metrics: We evaluate our approach against two baselines: time-domain backprojection and Reed et al. [43] across three dataset types (Simulated, AirSAS, and SVSS). Evaluation utilizes four core metrics—Chamfer distance, intersection-over-union (IoU), precision, and F1-score—on both point cloud and mesh representations. For point cloud metrics, we compare the model-generated point cloud against a reference point cloud derived from the ground truth mesh. For mesh-based metrics, we convert the generated point cloud to a mesh and then back to a point cloud, enabling a consistency check against the ground truth mesh's point cloud. Importantly, metric scores differ between raw point clouds and those produced through mesh conversion.

6. Experimental Results

Simulated Results: We evaluate our method on four simulated scenes and compare against backprojection and Reed et al. [43]. All scenes use a linear frequency-modulated (LFM) chirp with center frequency $f_c = 20\,\text{kHz}$ and bandwidth $\Delta f = 20\,\text{kHz}$ at an SNR of $20\,\text{dB}$. Fig. 2 shows qualitative results for Buddha and Armadillo. Tab. 1 reports 3D metrics averaged across all scenes, where our method achieves the best overall performance. Per-scene results and additional metrics are provided in the supplemental material.

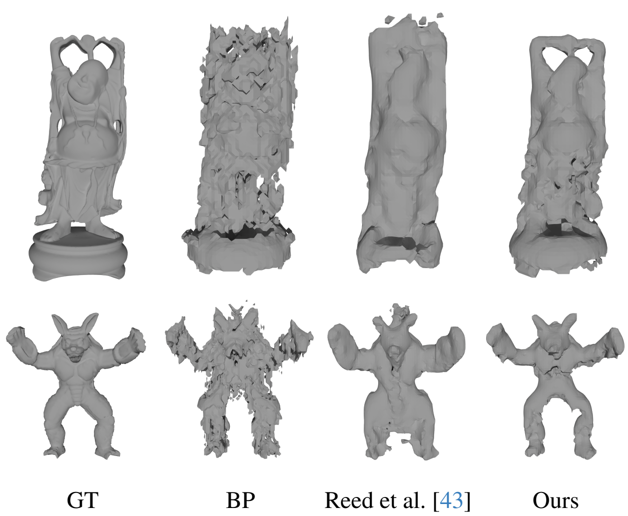

Figure 2. Simulated results for Buddha (top row) and Armadillo (bottom row) at 20 dB SNR.

AirSAS Results: We evaluate on real AirSAS measurements of the 3D-printed *Armadillo* and *Bunny*. All acquisitions use an LFM chirp with $f_c = 20\,\text{kHz}$ and $\Delta f = 20\,\text{kHz}$. Fig. 3 compare backprojection, Reed et al. [43], and our method.

Table 1. Average 3D metrics across all objects/methods. Chamfer in scientific notation; others rounded to 3 decimals.

	Method	Chamfer (↓)	IoU (↑)	Prec. (↑)	F1 (↑)
Point cloud	Backproj.	3.04e-4	0.201	0.244	0.333
	Reed et al.	1.96e-4	0.447	0.428	0.566
	Ours	**8.89e-5**	**0.497**	**0.545**	**0.616**
Mesh	Backproj.	3.85e-4	0.161	0.225	0.272
	Reed et al.	3.97e-4	0.133	0.171	0.232
	Ours	**9.70e-5**	**0.240**	**0.395**	**0.384**

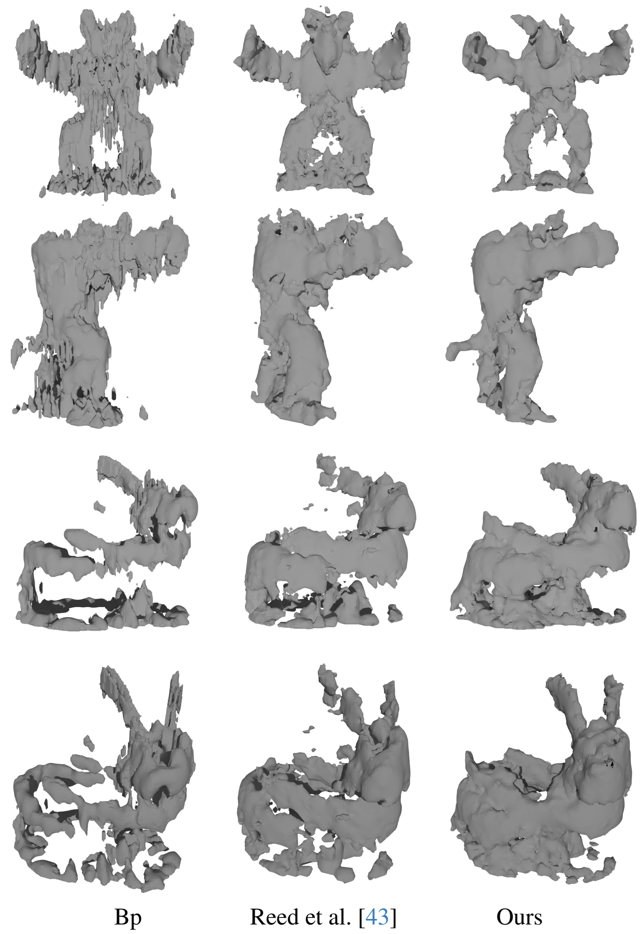

Figure 3. Real AirSAS results for Armadilo and Bunny at $\Delta f = 20\,\text{kHz}$

Qualitatively, our reconstructions align best with the ground truth: the legs and tail are better preserved and less fragmented, and radial streaking artifacts are reduced relative to backprojection. Compared to Reed et al., our method also yields sharper surfaces with fewer fill-in artifacts, especially in side and front views.

Subsampling the Synthetic Aperture Views: In addition to fully sampled reconstructions, we evaluate our method under sparse view sampling, where only 20% of the available views are used. As expected, backprojection

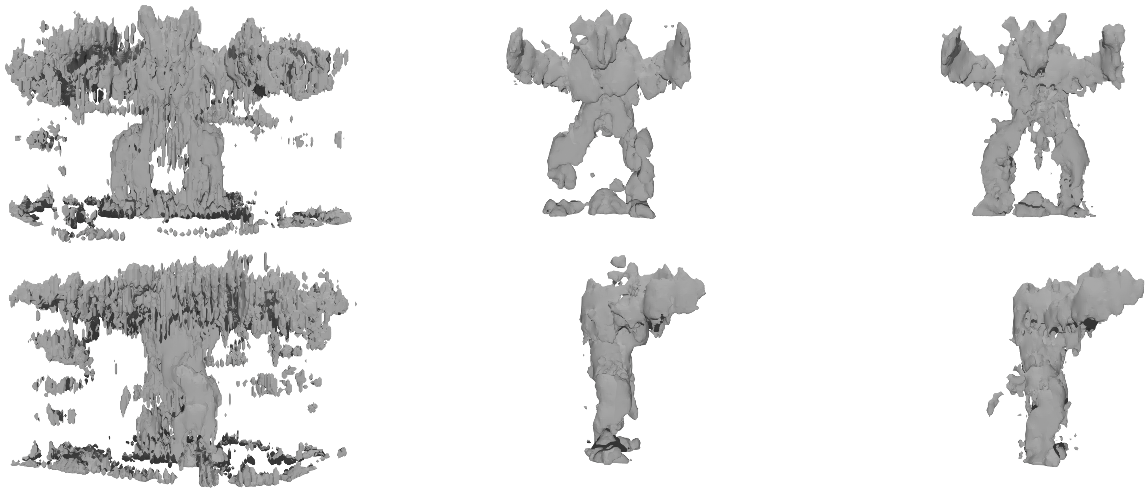

Figure 4. Sparse subsampling with 20% of views. Top: front view. Bottom: side view.

Table 2. Training iterations and wall-clock time on an RTX 4090.

Dataset	Method	Iterations	Time
Simulated	Reed et al. [43]	100,000	2 h 10 min
	Ours	50,000	44 min
AirSAS	Reed et al. [43]	26,000	1 h 38 min
	Ours	26,000	1 h 15 min
SVSS	Reed et al. [43]	6,000	38 min
	Ours	6,000	30 min

under sparse sampling exhibits strong radial streaking artifacts due to missing angular information. In contrast, our method produces reconstructions that are both sharper and more faithful to the ground truth. Notably, Fig. 4 shows that our method outperform both backprojection and the approach of Reed et al. [43] on the tested dataset, demonstrating that our representation can robustly handle limited angular coverage.

Latency and Convergence: All timings were measured on an RTX 4090 GPU; see Tab. 2 for dataset-level iteration counts and wall-clock times. Our *per-iteration* latency—measured end-to-end, including surface-normal computation, transmission-probability accumulation, and ray integration—averages $140\,\mathrm{ms}$ for Reed et al. [43] versus $125\,\mathrm{ms}$ for ours.

Ablation Studies: We provide ablations and additional results in the supplemental material. First, we include additional simulated results with qualitative comparisons and per-object views. Second, an SH-level ablation varies the maximum degree $L \in \{1,2,3\}$ and reports qualitative trends; we omit $L{=}0$ since it matches the isotropic baseline [43] in the main text. Third, a threshold sensitivity study compares iso-surface thresholds for marching cubes to illustrate reconstruction stability. Finally, signal fitting and novel-view synthesis present ToF fitting curves on training baselines and held-out TX–RX pairs to demonstrate generalization.

Underwater Reconstruction from SVSS: The SVSS scene contains two targets, a cylindrical pipe and a cinder block, placed on the lakebed (Fig. 5, right). Because the acquisition is done overhead on the lake surface by the sonar, only the near, exposed surfaces are insonified, while the undersides and much of the far side remain in acoustic shadow. A physically consistent reconstruction should therefore show a flattened side profile and a footprint consistent with the visible faces, with little to no structure below the object. In Fig. 5, backprojection produces rounded sidewalls with streaking and clear spurious energy beneath both targets, and Reed et al. [43] reduces but does not fully remove this curvature and bleed-through. Our method yields flatter side views and cleaner, more compact footprints for both the pipe and the cinder block (dashed boxes), while strongly suppressing returns below the objects, consistent with the one-sided visibility imposed by the SVSS geometry.

7. Discussion

To our knowledge, this is the first SAS reconstruction method that parameterizes the *complex* acoustic scattering field with spherical harmonics (SH) inside an implicit neural representation trained directly from 1D ToF signals. Across simulated, in-air (AirSAS), and underwater (SVSS) experiments, modeling view dependence via higher-order SH consistently outperforms isotropic INR baselines (e.g.,

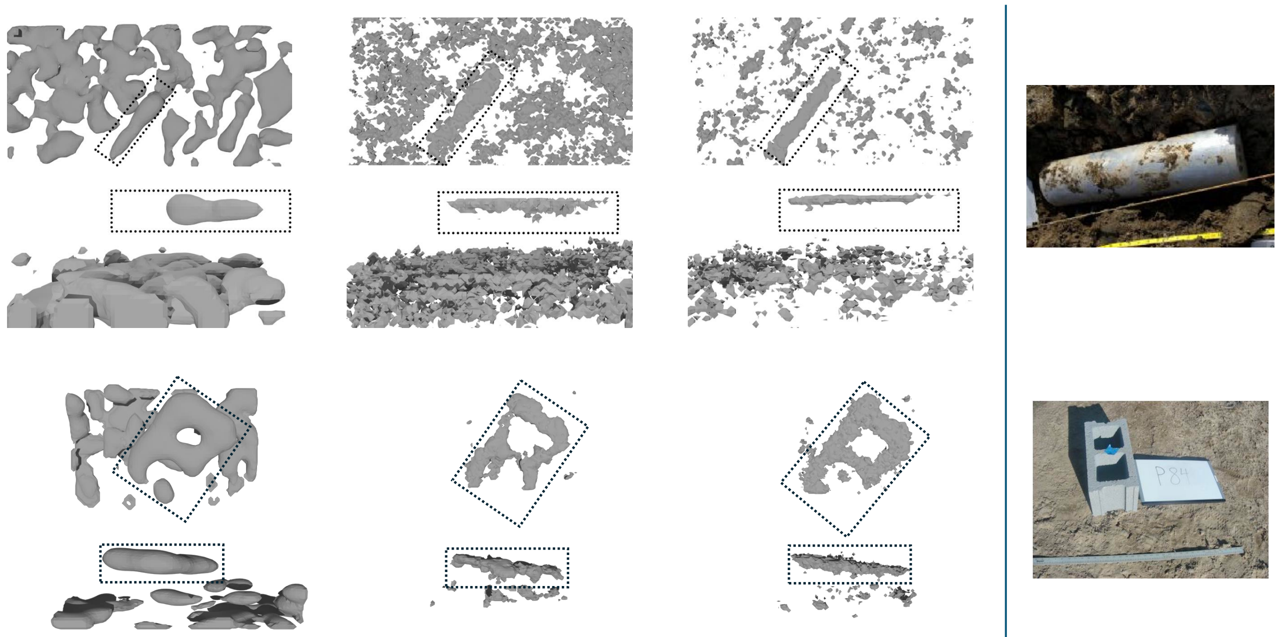

Figure 5. Underwater reconstruction from SVSS. Compared to backprojection and Reed et al. [43], our method recovers the expected flat side profile and rectangular footprint of the cylindrical pipe, while suppressing spurious returns below the target.

[43]). The DC term provides a stable isotropic density proxy, while higher orders compactly capture directional scattering—yielding sharper geometry under sparse views and fewer artifacts (e.g., better leg/tail preservation on real objects and a flatter side profile with rectangular footprint on the SVSS cylinder, consistent with one-sided visibility).

Our approach also converges faster in practice than isotropic INR while maintaining similar per-iteration cost (Sec. 6). These results suggest that directional parameterizations are a better inductive bias for SAS than purely isotropic scattering.

Limitations: Our approach assumes moderate SNR and sufficient bandwidth. In very noisy acquisitions, the higher-order SH terms ($\ell > 0$) act like high-frequency angular basis functions that readily fit noise, introducing spurious anisotropy and ripples that corrupt the geometry. With narrowband signals the problem worsens: range resolution degrades as $\Delta R \approx c/(2\Delta f)$ and pulse-compression gain scales with the time–bandwidth product, so returns from neighboring depths blur together and the directional components become weakly identifiable, making the SH fit ill-conditioned. Finally, our method does not model complex acoustic wave phenomena (e.g., diffraction, interference, and multiple reflections/scattering); incorporating these effects would require differentiable forward models integrated into our framework and is left for future work.

Future Work: Beyond SAS, we plan to test this technique on other time-of-flight modalities—particularly radar and LiDAR—by swapping in modality-specific forward models (e.g., wavelength and antenna/polarization effects for radar; visibility/BRDF terms for LiDAR). We hypothesize that the SH-based view-dependent modeling that improves SAS will likewise benefit these sensors. In addition, we plan to explore 3D diffusion priors as learned regularizers: beyond sparsity, phase smoothness, and bounded density, a diffusion prior could suppress artifacts under sparse views. Because public SVSS/AirSAS data are limited, one might be able to pre-train such priors on large-scale synthetic data from our simulator and then adapt to real data with lightweight fine-tuning.

8. Acknowledgement

This research was supported in part by ONR grant N00014-23-1-2406, NSF CCF-2326904, NSF CCF-2326905, and a gift by RTX, Inc. GPU support for this research was provided by ASU Research Computing through the SOL supercomputer [25].

References

[1] Benjamin Attal, Eliot Laidlaw, Aaron Gokaslan, Changil Kim, Christian Richardt, James Tompkin, and Matthew O'Toole. Törf: Time-of-flight radiance fields for dynamic scene view synthesis. *Advances in Neural Information Processing Systems*, 34, 2021. 2

[2] Francesco Baralli, Michel Couillard, Jesus Ortiz, and Darwin G Caldwell. Gpu-based real-time synthetic aperture sonar processing on-board autonomous underwater vehicles. In *2013 MTS/IEEE OCEANS-Bergen*, pages 1–8. IEEE, 2013. 2

[3] Emile Barbier-Renard, Florence Tupin, Nicolas Trouvé, and Loïc Denis. Multi-view 3d surface reconstruction from sar images by inverse rendering. *arXiv preprint arXiv:2502.10492*, 2025. 2

[4] Andrea Bellettini and Marc Pinto. Design and experimental results of a 300-khz synthetic aperture sonar optimized for shallow-water operations. *IEEE Journal of Oceanic Engineering*, 34(3):285–293, 2008. 1

[5] Thomas E. Blanford, John D. McKay, Daniel C. Brown, Joonho D. Park, and Shawn F. Johnson. Development of an in-air circular synthetic aperture sonar system as an educational tool. *Proceedings of Meetings on Acoustics*, 36(1): 070002, 2019. 5

[6] Thomas E Blanford, Luke Garrett, J Daniel Park, and Daniel C Brown. Leveraging audio hardware for underwater acoustics experiments. In *Proceedings of Meetings on Acoustics 182ASA*, page 030002. Acoustical Society of America, 2022. 5

[7] Thomas E Blanford, David P Williams, J Daniel Park, Brian T Reinhardt, Kyle S Dalton, Shawn F Johnson, and Daniel C Brown. An in-air synthetic aperture sonar dataset of target scattering in environments of varying complexity. *Scientific Data*, 11(1):1196, 2024. 1

[8] D. Borts, E. Liang, T. Broedermann, A. Ramazzina, S. Walz, E. Palladin, J. Sun, D. Brueggemann, C. Sakaridis, L. Van Gool, M. Bijelic, and F. Heide. Radar fields: Frequency-space neural scene representations for fmcw radar. In *Proc. SIGGRAPH*, 2024. 2

[9] Daniel C Brown. *Modeling and measurement of spatial coherence for normal incidence seafloor scattering*. PhD thesis, The Pennsylvania State University, 2017. 2

[10] Daniel C Brown, Shawn F Johnson, and Derek R Olson. A point-based scattering model for the incoherent component of the scattered field. *The Journal of the Acoustical Society of America*, 141(3):EL210–EL215, 2017. 1, 2

[11] Daniel C Brown, Shawn Johnson, Caleb Brownstead, Joseph Calantoni, and Gene Fabian. Sediment volume search sonar development. Technical report, SERDP/ESTCP, 2021. `https://serdp-estcp.org/projects/details/97e4a43c-bb65-432d-94ba-6aaff361571a`. 5

[12] Hayden J Callow. *Signal Processing for Synthetic Aperture Sonar Image Enhancement*. PhD thesis, University of Canterbury, 2003. 1, 2, 5

[13] Hongbo Chen, Logiraj Kumaralingam, Shuhang Zhang, Sheng Song, Fayi Zhang, Haibin Zhang, Thanh-Tu Pham, Kumaradevan Punithakumar, Edmond HM Lou, Yuyao Zhang, et al. Neural implicit surface reconstruction of freehand 3d ultrasound volume with geometric constraints. *Medical Image Analysis*, 98:103305, 2024. 2

[14] Benjamin Cowen, J Daniel Park, Thomas E Blanford, Geoff Goehle, and Daniel C Brown. Airsas: Controlled dataset generation for physics-informed machine learning. In *NeurIPS Data-Centric AI Workshop*, 2021. 5

[15] Junyuan Deng, Pengfei Xie, Lei Zhang, and Yulai Cong. Isar-nerf: Neural radiance fields for 3-d imaging of space target from multiview isar images. *IEEE Sensors Journal*, 24(7):11705–11722, 2024. 2

[16] Jerry Eaves and Edward Reedy. *Principles of modern radar*. Springer Science & Business Media, 2012. 2

[17] Rudiger Gens and John L Van Genderen. Review article sar interferometry—issues, techniques, applications. *International Journal of Remote Sensing*, 17(10):1803–1835, 1996. 2

[18] Isaac D Gerg, Daniel C Brown, Stephen G Wagner, Daniel Cook, Brian N O'Donnell, Thomas Benson, and Thomas C Montgomery. Gpu acceleration for synthetic aperture sonar image reconstruction. In *Global Oceans 2020: Singapore–US Gulf Coast*, pages 1–9. IEEE, 2020. 1, 2

[19] Geoff Goehle, Benjamin Cowen, Thomas E. Blanford, J. Daniel Park, and Daniel C. Brown. Approximate extraction of late-time returns via morphological component analysis. *The Journal of the Acoustical Society of America*, 153 (5):2838, 2023. 5

[20] HD Griffiths, TA Rafik, Z Meng, CFN Cowan, H Shafeeu, and DK Anthony. Interferometric synthetic aperture sonar for high-resolution 3-d mapping of the seabed. *IEE Proceedings-Radar, Sonar and Navigation*, 144(2):96–103, 1997. 2

[21] Roy Edgar Hansen, Torstien Olsom Saebo, Kenneth Gade, and Sean Chapman. Signal processing for auv based interferometric synthetic aperture sonar. *IEEE Oceans 2003*, 5: 2438–2444, 2003. 2

[22] Roy Edgar Hansen, Hayden John Callow, Torstein Olsmo Sabo, and Stig Asle Vaksvik Synnes. Challenges in seafloor imaging and mapping with synthetic aperture sonar. *IEEE Transactions on Geoscience and Remote Sensing*, 49(10): 3677–3687, 2011. 1

[23] Michael P Hayes and Peter T Gough. Synthetic aperture sonar: A review of current status. *IEEE Journal of Oceanic Engineering*, 34(3):207–224, 2009. 1, 2

[24] Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, and Anthony Rowe. Dart: Implicit doppler tomography for radar novel view synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 24118–24129, 2024. 2

[25] Douglas M. Jennewein, Johnathan Lee, Chris Kurtz, William Dizon, Ian Shaeffer, Alan Chapman, Alejandro Chiquete, Josh Burks, Amber Carlson, Natalie Mason, Arhat Kobawala, Thirugnanam Jagadeesan, Praful Bhargav Basani, Torey Battelle, Rebecca Belshe, Deb McCaffrey, Marisa Brazil, Chaitanya Inumella, Kirby Kuznia, Jade Buzinski, Dhruvil Deepakbhai Shah, Sean M. Dudley, Gil Speyer, and

Jason Yalim. The sol supercomputer at arizona state university. In *Practice and Experience in Advanced Research Computing 2023: Computing for the Common Good*, page 296–301, New York, NY, USA, 2023. Association for Computing Machinery. 8

[26] Josiah Jideani and Andrew Wilkinson. Air-based synthetic aperture sonar tomography using compressive sensing. In *2014 IEEE International Ultrasonics Symposium*, pages 49–52. IEEE, 2014. 1

[27] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, 2015. 5

[28] William E. Lorensen and Harvey E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. In *Proceedings of SIGGRAPH '87*, pages 163–169, 1987. 5

[29] Weihan Luo, Anagh Malik, and David B Lindell. Transientangelo: Few-viewpoint surface reconstruction using single-photon lidar. In *Proceedings of the Winter Conference on Applications of Computer Vision (WACV)*, pages 8712–8722, 2025. 2

[30] A. Malik, P. Mirdehghan, S. Nousias, K. N. Kutulakos, and D. B. Lindell. Transient neural radiance fields for lidar view synthesis and 3d reconstruction. In *Proc. NeurIPS*, 2023. 2

[31] Timothy M Marston, Jermaine L Kennedy, and Philip L Marston. Coherent and semi-coherent processing of limited-aperture circular synthetic aperture (csas) data. In *OCEANS'11 MTS/IEEE KONA*, pages 1–6. IEEE, 2011. 2

[32] Timothy M Marston, Jermaine L Kennedy, and Philip L Marston. Autofocusing circular synthetic aperture sonar imagery using phase corrections modeled as generalized cones. *The Journal of the Acoustical Society of America*, 136(2): 614–622, 2014. 2

[33] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Representing scenes as neural radiance fields for view synthesis. In *ECCV*, 2020. 2, 4

[34] Claus Müller. *Spherical harmonics*. Springer, 2006. 4

[35] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. *ACM TOG*, 41(4):1–15, 2022. 3

[36] Md Farhan Tasnim Oshim, Albert Reed, Suren Jayasuriya, and Tauhidur Rahman. Nerf-enabled analysis-through-synthesis for isar imaging of small everyday objects with sparse and noisy uwb radar data. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 4586–4593. IEEE, 2024. 2

[37] J Daniel Park, Thomas E Blanford, Daniel C Brown, and Daniel Plotnick. Alternative representations and object classification of circular synthetic aperture in-air acoustic data. *The Journal of the Acoustical Society of America*, 148(4): 2661–2661, 2020. 5

[38] Daniel S Plotnick, Philip L Marston, and Timothy M Marston. Fast nearfield to farfield conversion algorithm for circular synthetic aperture sonar. *The Journal of the Acoustical Society of America*, 136(2):EL61–EL66, 2014. 2

[39] Mohamad Qadri, Michael Kaess, and Ioannis Gkioulekas. Neural implicit surface reconstruction using imaging sonar. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 1040–1047. IEEE, 2023. 2

[40] Mohamad Qadri, Kevin Zhang, Akshay Hinduja, Michael Kaess, Adithya Pediredla, and Christopher A Metzler. Aoneus: A neural rendering framework for acoustic-optical sensor fusion. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–12, 2024. 2

[41] Ravi Ramamoorthi and Pat Hanrahan. A signal-processing framework for inverse rendering. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pages 117–128, 2001. 4

[42] Albert Reed, Thomas Blanford, Daniel C. Brown, and Suren Jayasuriya. Sinr: Deconvolving circular sas images using implicit neural representations. *IEEE Journal of Selected Topics in Signal Processing*, pages 1–16, 2022. 2

[43] A. Reed, J. Kim, T. Blanford, A. Pediredla, D. Brown, and S. Jayasuriya. Neural volumetric reconstruction for coherent synthetic aperture sonar. *ACM Trans. Graph.*, 42(4), 2023. 1, 2, 3, 4, 5, 6, 7, 8

[44] Hiroshi Saruwatari and Mitsuo Komura. Synthetic aperture sonar in air medium using a nonlinear sidelobe canceler. *Electronics and Communications in Japan (Part III: Fundamental Electronic Science)*, 82(12):13–25, 1999. 1

[45] Robert T Seeley. Spherical harmonics. *The American Mathematical Monthly*, 73(4P2):115–121, 1966. 4

[46] Siyuan Shen, Zi Wang, Ping Liu, Zhengqing Pan, Ruiqian Li, Tian Gao, Shiying Li, and Jingyi Yu. Non-line-of-sight imaging via neural transient fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2257–2268, 2021. 2

[47] Xiaoshen Song and Weidong Yu. Processing video-sar data with the fast backprojection method. *IEEE Transactions on Aerospace and Electronic Systems*, 52(6):2838–2848, 2016. 2

[48] Mehrdad Soumekh. *Synthetic aperture radar signal processing*. New York: Wiley, 1999. 2

[49] Harshvardhan Takawale and Nirupam Roy. Spinr: Neural volumetric reconstruction for fmcw radars. *arXiv preprint arXiv:2503.23313*, 2025. 2

[50] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. In *Proceedings of the 35th International Conference on Neural Information Processing Systems*, Red Hook, NY, USA, 2021. Curran Associates Inc. 4

[51] Yiping Xie, Giancarlo Troni, Nils Bore, and John Folkesson. Bathymetric surveying with imaging sonar using neural volume rendering. *IEEE Robotics and Automation Letters*, 9(9): 8146–8153, 2024. 2