## A    REPRODUCIBILITY

Anonymized code and demo datasets will be available on our webpage (https://anon61823-star.github.io). We provide details about comparison with other algorithms to facilitate the reproducing of our results. All details about the hyperparameters, environment specifications, and real-world experiment setup are provided in the appendix or the website.

## B    DERIVATION OF SPECTRAL TRANSFORMATION FOR FORWARD MODEL

To clarify the mathematical foundation of our *spectral forward model* in SpINRv2 (Ours), we revisit the discrete Fourier analysis of a single complex tone. This explicit derivation highlights two properties that are crucial for radar reconstruction: (i) how a finite-length DFT distributes energy into neighbouring bins when a scatterer's beat frequency is *off-bin*, and (ii) why modelling both magnitude *and* phase is essential for physically faithful supervision.

**Signal model.**    We consider the sampled beat signal from one ideal point scatterer,

$$S_t = M\, e^{\mathrm{i}(\alpha t + \phi)}, \qquad \alpha = \frac{2\pi f}{N}, \quad t = 0, \dots, N-1, \tag{1}$$

where $M$ is the amplitude, $\phi$ the initial phase, $f$ the (normalised) frequency and $N$ the analysis window length.

**DFT definition.**    The unit-normalised DFT coefficient at bin index $k$ is

$$Z_k = \frac{1}{N} \sum_{t=0}^{N-1} S_t\, e^{-\mathrm{i}\beta_k t}, \qquad \beta_k = \frac{2\pi k}{N}. \tag{2}$$

**Forming a geometric series.**    Substituting $S_t$ and collecting constants gives

$$Z_k = \frac{M}{N}\, e^{\mathrm{i}\phi} \sum_{t=0}^{N-1} \big(e^{\mathrm{i}(\alpha - \beta_k)}\big)^t, \tag{3}$$

which is a finite geometric series with common ratio $r = e^{\mathrm{i}(\alpha - \beta_k)}$.

**Closed-form sum.**    Using the standard sum formula $\sum_{t=0}^{N-1} r^t = (1 - r^N)/(1 - r)$ we obtain

$$Z_k = \frac{M}{N}\, e^{\mathrm{i}\phi}\, \frac{1 - e^{\mathrm{i}(\alpha - \beta_k)N}}{1 - e^{\mathrm{i}(\alpha - \beta_k)}}. \tag{4}$$

**Eliminating the bin phase.**    Because $\beta_k N = 2\pi k$ is an integer multiple of $2\pi$, $e^{-\mathrm{i}\beta_k N} = 1$ and equation 4 simplifies to the closed form used in the main paper:

$$Z_k = \boxed{\frac{M}{N}\, e^{\mathrm{i}\phi}\, \frac{1 - e^{\mathrm{i}\alpha N}}{1 - e^{\mathrm{i}(\alpha - \beta_k)}}}. \tag{2 revisited}$$

**Interpretation.**    If the scatterer frequency *exactly* matches a bin centre ($\alpha = \beta_k$), both numerator and denominator of equation 2 revisited vanish; evaluating the limit (or returning to the original sum) yields $Z_k = M\, e^{\mathrm{i}\phi}$ for that bin and $Z_{k'} = 0$ elsewhere. For all *off-bin* frequencies ($\alpha \neq \beta_k$), equation 2 revisited predicts a sinc-shaped spectral envelope—the familiar *spectral leakage* modelled in SpINR. This can also be seen in Fig. 1 and 2. Capturing this leakage, including complex phase rotations, is what allows our frequency-domain supervision to remain physically consistent even when a scatterer's true delay does not align with a DFT bin centre.

## C    ADDITIONAL EVALUATION

### C.1    ADDITIONAL QUALITATIVE RESULTS FOR RECONSTRUCTION

In this section we show some additional results for scene reconstruction with SpINRv2 (Ours) and othe baselines.
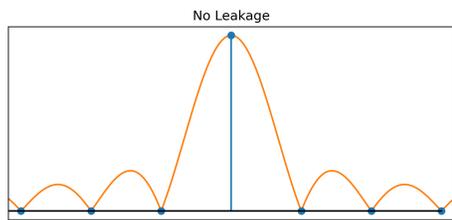
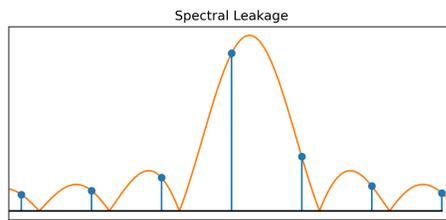Figure 1: When the $\alpha = \beta_k$, there is no leakage



Figure 2: When the $\alpha \neq \beta_k$, there is leakage and in neighbouring bins

| Range Quant. | TF-TS | TF-SS | Backprojection | SpINR (Ours) | Ground Truth |
|---|---|---|---|---|---|



Figure 3: Comparison of volumetric reconstructions

## C.2  COMPUTATION EFFICIENCY OF SPINRV2 (OURS) VS. TF-TS VS. RQ

Fig. 4 shows us a comparison between computation efficiency and quality of reconstruction for various models. SpINRv2 (Ours) outperforms Time domain forward in both runtime efficiency and chamfer distance. Range Quantization is much faster than SpINRv2 (Ours)'s forward model but severely underperforms in reconstruction quality.

## C.3  RECONSTRUCTION QUALITY OF SPINRV2 (OURS) VS. CLASSICAL COHERENT BACKPROJECTION

To further validate the effectiveness of our proposed frequency-domain forward model, we compare it against a classical reconstruction baseline: coherent backprojection (CBP). This method discretizes the scene into voxels and accumulates backprojected energy from all Tx paths by aligning phases based on path delays. While physically grounded, CBP has significant limitations that our learning-based approach addresses.

(1) Inherent Discretization and Aliasing: CBP relies on uniform voxel grids and discrete binning of round-trip delays, which often leads to aliasing artifacts and blurring, especially when the sampling
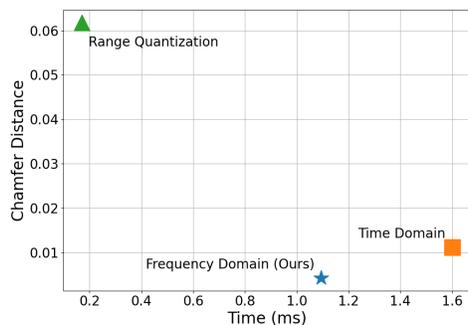
Figure 4: We compare the three forward models for their runtime and volume reconstruction accuracy. Even though the Range Quantization has the fastest runtime, it severe underperforms in volume reconstruction as signified by the chamfer distance (lower better).

aperture is non-uniform or sparse. Our method, by contrast, uses a continuous volumetric representation (via neural fields) and avoids explicit voxelization, enabling sub-voxel accuracy in both representation and rendering.

(2) Lack of Differentiability and Learning: CBP is a purely geometric method—it does not learn from data or optimize a forward model. As a result, it is highly sensitive to aperture coverage, noise, and partial views. Our method incorporates a fully differentiable, physics-informed forward model, enabling gradient-based optimization that refines the scene representation to best match all measurements.

(3) Better Use of Redundant Views: In cases with dense or overlapping Tx-Rx observations, CBP merely accumulates energy, often leading to oversmoothened reconstructions. In contrast, our method learns to fuse redundant observations through a shared implicit scene representation, enabling better consistency across viewpoints and preservation of fine geometry.

As shown in our quantitative results (Tables 1,and 2, SpINR achieves significantly better reconstruction compared to CBP.

In summary, while CBP serves as a useful classical baseline, it lacks the flexibility, expressiveness, and robustness offered by our differentiable frequency-domain formulation paired with neural reconstruction. This synergy between physics-based modeling and data-driven learning is key to achieving high-fidelity 3D volumetric reconstructions from FMCW radar measurements.

## C.4 EFFECT OF BANDWIDTH CHANGE ON RECONSTRUCTION QUALITY

The spatial resolution of FMCW radar systems is fundamentally limited by the chirp bandwidth $B$, with the minimum distinguishable range given by $\frac{c}{2B}$. As the bandwidth decreases, this resolution limit coarsens, which traditionally leads to severe blurring in reconstruction techniques such as backprojection. However, SpINR is not bound by explicit voxel grids or range quantization. Instead, it learns a continuous implicit representation of the scatterer field. This allows the network to infer and preserve spatial structure even under low-bandwidth conditions, where conventional methods lose detail.

Figure 5 compares reconstructions from SpINR and classical backprojection as the bandwidth is reduced from 4 GHz down to 40 MHz—a 100x decrease. While both methods suffer some degradation in spatial sharpness, SpINR exhibits a notably graceful degradation. Key object features remain preserved at lower bandwidths, and the reconstructions retain their overall shape and topology. In contrast, backprojection rapidly collapses into severely blurred and aliased outputs. These results highlight the robustness of the learned representation in SpINR to physical resolution limits imposed by the sensor, demonstrating its utility in bandwidth-constrained regimes.
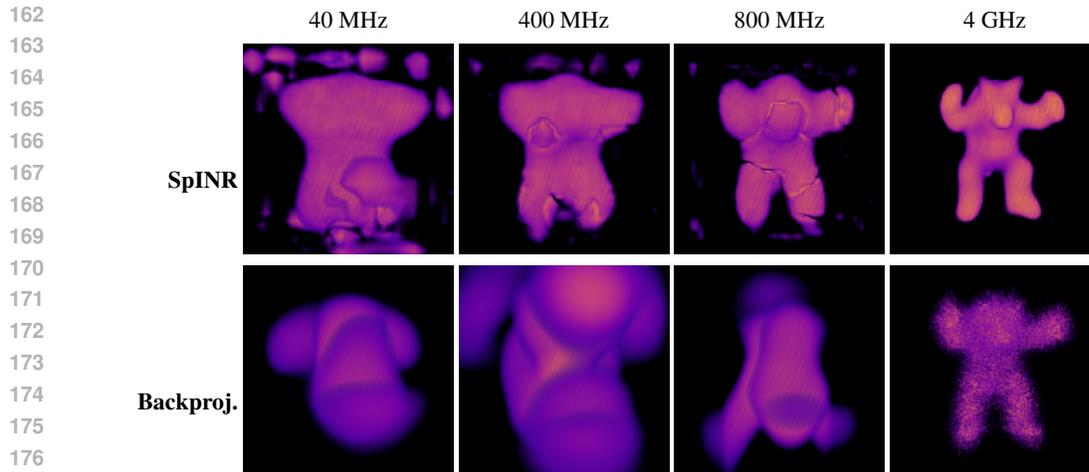
3

Figure 5: Comparison of volumetric reconstructions for different bandwidths (4GHz, 800 MHz, 400 MHz, 40 MHz) using SpINR and classical backprojection. SpINR shows a graceful degradation as compared with backprojection.

**Effect of Start Frequency on Reconstruction.**   Table 1 highlights how the choice of start frequency $f_0$ influences reconstruction quality in the absence of regularization. Although range resolution is determined by the signal bandwidth, increasing $f_0$ leads to shorter wavelengths and thus more severe phase wrapping—making it difficult to resolve scatterers at sub-bin offsets. This effect appears as a multiplicative phase term $e^{i\phi}$, where $\phi = 2\pi f_0 \tau$. As shown in the Fig. **??**, as the frequencies increase, the loss landscape becomes non-flat and the ambiguity increases. These observations further motivate the use of regularization to mitigate ambiguities arising from off-bin scatterer positions.
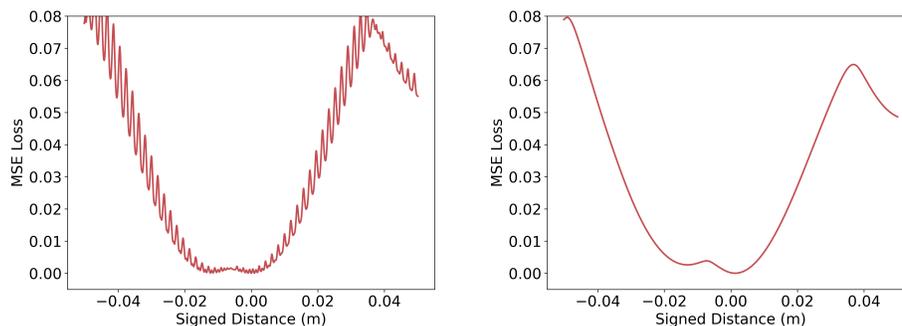


Figure 6: Loss when $\lambda/4 < \Delta r$ vs. loss when $\lambda/4 > \Delta r$. As the start frequency increases the loss landscaped becomes non-flat due to ambiguity.

| Frequencies | IoU ($\uparrow$) | Chamfer ($\downarrow$) | Hausdorff ($\downarrow$) | PSNR ($\uparrow$) | SSIM ($\uparrow$) | LPIPS ($\downarrow$) |
|---|---|---|---|---|---|---|
| 0 | 0.1312 | 0.0050 | 0.0507 | 16.6148 | 0.6971 | 0.2799 |
| 100 | 0.1311 | 0.0050 | 0.0507 | 16.6203 | 0.6974 | 0.2793 |
| 1e6 | 0.1303 | 0.0050 | 0.0507 | 16.5800 | 0.6950 | 0.2849 |
| 100e6 | 0.1328 | 0.0049 | 0.0507 | 16.7841 | 0.6960 | 0.2825 |
| 500e6 | 0.1243 | 0.0054 | 0.1179 | 14.8324 | 0.5537 | 0.5336 |
| 1e9 | 0.0313 | 0.0278 | 0.1514 | 7.2550 | 0.3479 | 0.7283 |

Table 1: Effect of Start frequency on Reconstruction without regularization

4

## C.5 Evaluation on a High-Frequency Sheet Benchmark

To evaluate the spatial resolution limits of SpINR, we introduce a synthetic benchmark object designed to gradually vary in surface detail. Specifically, we construct a vertical sheet where sinusoidal undulations increase in frequency along the y-axis (Fig. 7). The initial part of the object features smooth, low-frequency ripples, while the latter part contains increasingly fine patterns that challenge the system's resolving capacity.

We render volumetric reconstructions using SpINR and compare the learned scatterer field with the ground-truth surface. As shown in Fig. 8, SpINR accurately captures coarse structures in the first half, with clear alignment between the predicted maxima and the ground-truth scatterer locations. However, in the high-frequency regions, the system begins to miss fine oscillations, resulting in a loss of structural fidelity and localized blurring.
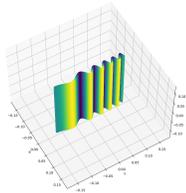


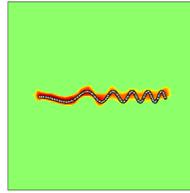Figure 7: (a) Ground-truth 3D sheet geometry

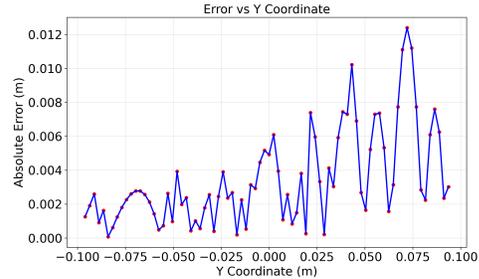Figure 8: (b) Overlay of predicted maxima (red) with GT (white)

Figure 9: (c) Error vs. vertical Y-axis

## C.6 Additional Results on Gradient Analysis

As explained in the main paper, We visualize the gradient statistics of both forward models (Figure 10). Our method (green) exhibits significantly more stable gradient statistics than the time-domain model (gray). The variance remains controlled throughout training, avoiding sudden spikes or vanishing behavior.

Our empirical observations parallel a broader body of work showing that architectural choices which preserve or align gradients promote both stable optimization and stronger generalization. In the context of FMCW radar signal modeling, we extend this principle by introducing a forward model rooted in spectral-domain physics. The resulting design not only enhances interpretability but also yields smoother, more effective gradient flow, thereby facilitating faster and more reliable training.

## D  Reproducibility

The code for dataset creation, model training and evaluation can be found in the "MMwave_recon" folder. The folder has a "README.md" which contains all the steps for running the code.

## E  Hyperparameters

270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323

(a) Mean for first layer

(b) Standard Deviation for last layer

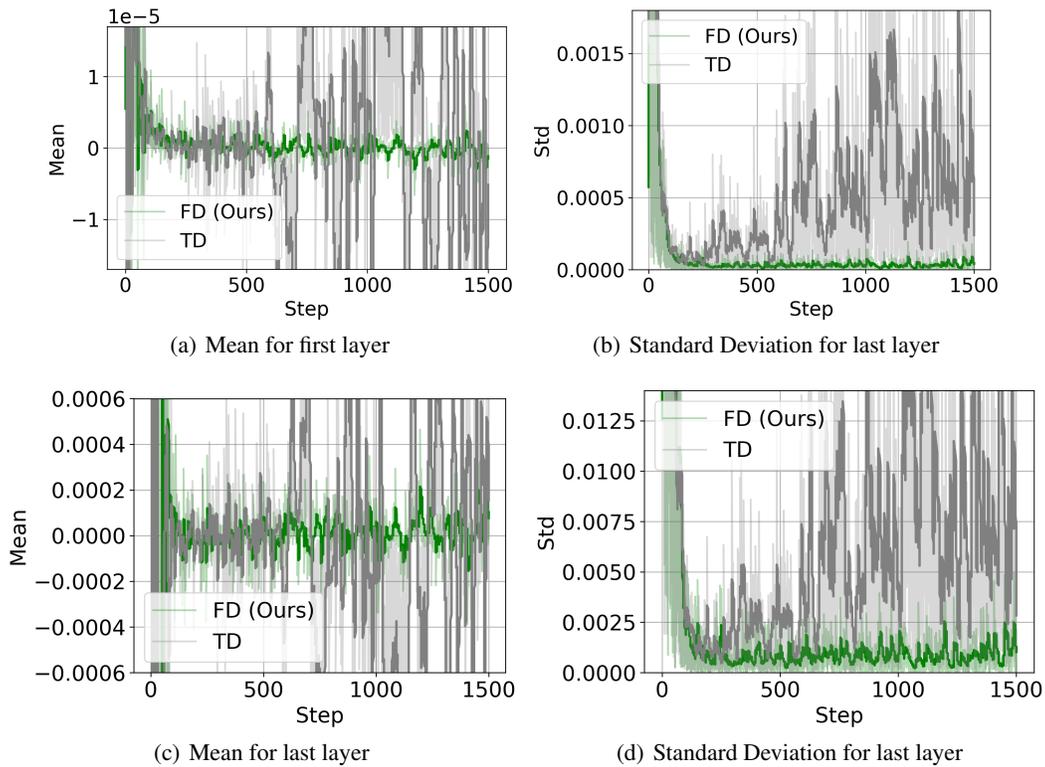(c) Mean for last layer

(d) Standard Deviation for last layer

Figure 10: We log the (a) mean and (b) standard deviation of the gradients for the first layer w.r.t. the number of training steps. The mean and standard deviation are more stable for our proposed method. For the time domain forward model the the gradient tends to explode. We observe a similar trend for all the subsequent layers.

324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377

| Parameter | Default Value | Description |
|---|---|---|
| **Learning & Optimization** | | |
| `learning_rate` | 1e-5 | Learning rate for network training |
| `num_epochs` | 345600 | Number of transmitter locations to train on |
| `accum_grad` | 1 | Gradient accumulation steps |
| `batch_size` | 1 | Training batch size |
| **Network Architecture** | | |
| `num_layers` | 12 | Number of hidden layers in the network |
| `num_neurons` | 256 | Number of neurons per hidden layer |
| **Loss Weights** | | |
| `scale_factor` | 0.00005 | Overall loss scaling factor |
| `mse_weight` | 0.01 | Weight for MSE loss component |
| `smooth_loss` | 1e4 | Weight for smoothness regularization |
| `sparsity` | 5e3 | Weight for sparsity regularization |
| **Occlusion & Geometry** | | |
| `occlusion_scale` | 1e0 | Scale factor for occlusion handling |
| `beamwidth` | 90 | Radar beam width in degrees |
| **Sampling & Rays** | | |
| `num_scatterers` | 40000 | Number of scatterer points sampled per iteration for random sampling |
| `num_rays` | 2200 | Number of rays processed per batch |
| `max_voxels` | 15000 | Maximum number of voxels to process |
| `sampling_distribution_uniformity` | 1.0 | Uniformity of sampling distribution |

Table 2: Default hyperparameters used in training and evaluation.