LEARNING 4D MATERIAL-INTERFACE DYNAMICS FROM FEW X-RAY PROJECTIONS

Nadav Schneider
Ben-Gurion University, IAEC
Israel
nadavsch@post.bgu.ac.il

Gershon Hanoch
Ben-Gurion University, NRCN
Israel
hanochge@post.bgu.ac.il

Gal Oren
Stanford University, Technion
United States
galoren@stanford.edu

ABSTRACT

Reconstructing the time-varying 3D material-interface state of a two-material system from only a few X-ray projections remains an unsolved challenge. Conventional tomography breaks down in this regime, and physics-driven trial-and-error is slow and biased. We present a learning-based framework for dynamic sparse-view tomography that addresses both data scarcity in real experiments and the ill-posedness of time-resolved 3D recovery. First, we introduce the *Synthetic Dynamic Tomography* (SDT) dataset: a scalable pipeline that lifts inexpensive 2D simulations of interfacial instability into 4D ground-truth volumes and physically realistic radiographs while varying noise, view sparsity, and symmetry breaking. Second, we adapt neural radiance fields to represent evolving 3D states as continuous 4D functions (3D space + time), using SDT to learn priors and enabling optimization from sparse projections when volume supervision is unavailable. On Rayleigh–Taylor benchmarks, our model reconstructs coherent spike–bubble dynamics from as few as one view per timestep and remains robust to moderate symmetry breaking. Code and data: https://github.com/Scientific-Computing-Lab/SDT.

1 INTRODUCTION

Reconstructing the time-varying 3D structure of a *two-material interface* from only a handful of X-ray projections is a fundamental challenge in dynamic experimental characterization. Many high-speed experiments – from interfacial instabilities to materials under shock – can only be probed by a few radiographic views of an evolving 3D state, due to limited diagnostic access and the rapid timescales involved (3; 2). A canonical example is the Rayleigh–Taylor instability, where acceleration drives an unstable interface between two materials (often fluids or plasmas of different density), producing interpenetration and a growing mixing layer that is challenging to observe in 3D from sparse radiographs. For example, in inertial confinement fusion (ICF) implosions, capturing more than one or two X-ray images during the microsecond event is extremely difficult, making full 3D computed tomography infeasible (10; 12). Classical reconstruction methods like filtered back-projection strictly require many angles, so they break down in this sparse-view, dynamic regime (19; 23). As a result, experimentalists have often relied on physics-based approaches: iteratively tuning detailed simulations until their 2D projections qualitatively match the measured radiographs. This manual "shot reconstruction" by trial-and-error is painstaking, biased, and explores only a tiny fraction of plausible 3D trajectories.

Recently, data-driven methods have shown promise in overcoming these limitations. Convolutional neural networks (CNNs) and neural rendering techniques can learn to infer 3D structure from very limited views, given enough training examples (17; 1). For instance, a CNN trained on synthetic

ICF capsule images was able to reconstruct the 3D shell morphology from a single radiograph (assuming the object is nearly symmetric) (18). Neural radiance fields (NeRFs), which represent a volume as a continuous 3D function, have also been applied to sparse-view medical imaging – e.g., MedNeRF can recover a full 3D CT image from just one or few X-rays by learning a continuous representation of anatomy (4; 20). These learning-based approaches greatly reduce the reliance on multiple projections, and NeRF models have recently been proposed as a paradigm shift to achieve 3D reconstructions during dynamic events (5). Early results demonstrate that neural radiance fields can outperform conventional limited-angle tomography methods on both synthetic and experimental data (22). However, a key hurdle for applying such models to rapidly-evolving experiments is the lack of large, labeled training datasets — one cannot obtain ground-truth 3D volumes for supervision, and real experiments yield only a few images per event.

In this work, we address the data scarcity problem and present a learning approach for time-resolved 3D reconstruction from extremely sparse X-ray views. We introduce a scalable pipeline to generate unlimited synthetic training data for dynamic tomography of interfacial mixing, and we adapt a neural radiance field model to learn a continuous 4D representation (3D space + time) of the evolving state (Figure 1). We demonstrate this approach on a challenging benchmark based on the Rayleigh–Taylor instability, and discuss how it generalizes to other dynamic two-material phenomena. Our contributions are: (i) a reproducible data-generation pipeline that converts previously inaccessible dynamic regimes into supervised training corpora, and (ii) a radiance-field baseline for sparse dynamic tomography, which we use to quantify current capabilities and pinpoint remaining gaps. By bridging physics simulations with modern machine learning, this work lays a foundation for data-driven inference of hidden 3D interface dynamics from minimal experimental imagery.

Implications for automated characterization under sparse diagnostics. While this work is motivated by fast experiments, the underlying setting – sparse-view radiography of a rapidly evolving two-material system – also arises in broader experimental characterization workflows where diagnostic access is constrained (e.g., limited-angle acquisition due to geometric constraints, or fast time-resolved CT that must operate with fewer projections) (21; 7). In such regimes, a learned 4D reconstruction can serve as an intermediate representation that reduces reliance on manual, simulation-tuning loops: instead of matching projections qualitatively, one can (when scientifically appropriate) extract quantitative descriptors directly from the reconstructed volume over time (e.g., interface geometry, mixing-region thickness, or other morphology statistics defined by the experimental community) (14; 8; 11). Importantly, we do not claim to solve downstream material property inference here; rather, we provide a controlled benchmark and a reproducible reconstruction baseline that can be integrated into future end-to-end pipelines for more automated analysis under extreme view sparsity.

2 Problem Setting and Motivation

In many fast, radiography-based experiments involving two-material systems (interfacial mixing, shocks, plasma implosions, etc.), we can record only a handful of 2D projection images of a transient 3D process. The goal is to recover the time-resolved 3D state of the system – and in particular the evolution of the material interface and mixing morphology – from these few projections at each time step. This problem is severely ill-posed: the object is evolving, imaging windows are brief, and there is no ground-truth 3D volume to directly supervise the reconstruction. Traditional CT scanning is too slow or radiation-intensive to capture such dynamics. Instead, practitioners have resorted to indirect means – for example, "digital volume correlation" techniques assume the object undergoes small deformations from a known reference volume, and can then solve for 3D displacement fields using only a few radiographs (13; 9). While powerful in certain regimes (e.g. solid mechanics), those methods break down when the object's topology or morphology changes dramatically (as in interfacial instabilities or ICF implosions). Another approach has been to fit parametric physics models to the projections: for instance, optimizing a low-dimensional model (e.g., a small set of parameters describing an interface or a material model) to match a sequence of radiographs over time. This can succeed for well-characterized systems, but it requires assuming a simplified model *a priori* and cannot capture unforeseen dynamics.

In short, existing physics-driven approaches either require strong prior knowledge (limited deformations, simple parametric forms) or an inordinate amount of manual trial-and-error. They are not

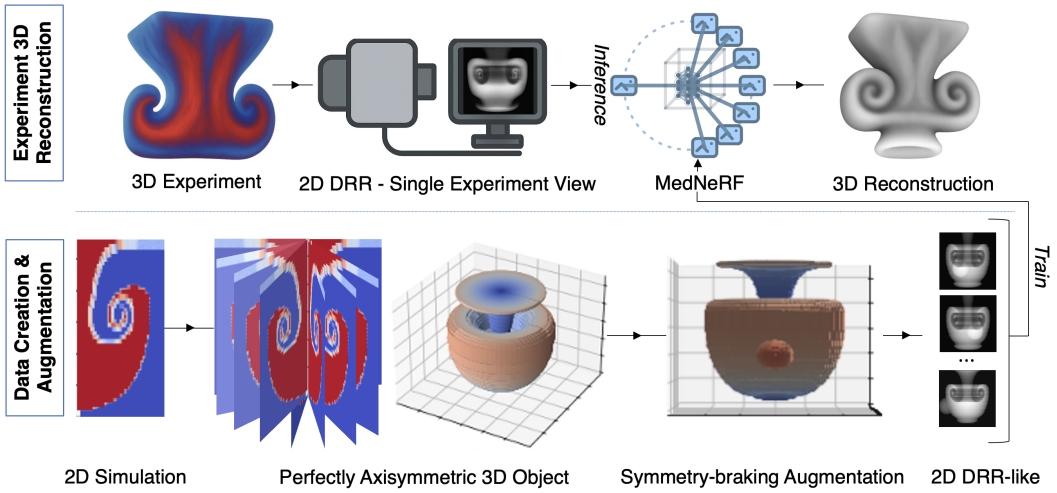


Figure 1: **Illustration of the proposed pipeline for dynamic sparse-view tomography of two-material interface dynamics.** (Top) In real experiments, a single 2D digitally reconstructed radiograph (DRR) of an evolving two-material system is used as input to a MedNeRF-based model, which infers a full 3D reconstruction despite the extreme view sparsity. (Bottom) To enable training, we generate synthetic data by lifting inexpensive 2D simulations of Rayleigh–Taylor interfacial instabilities into perfectly axisymmetric 3D volumes, followed by controlled symmetry-breaking augmentations. Forward-projected radiographs (DRR-like) from these 3D states provide virtually unlimited paired data for training. Together, this workflow bridges 2D simulations and sparse-view experimental data, enabling continuous 4D (3D+time) reconstruction of interface evolution under realistic conditions.

general solutions for arbitrary evolving 3D two-material phenomena under sparse imaging. This motivates a new, learning-based formulation: can we train a model to directly learn the mapping from a few projections to the full 3D state, by exposing it to many examples of physically plausible evolution? In particular, we ask whether such a system can reliably recover time-resolved 3D interface dynamics from very few views per timestep, how many views are actually needed, how sensitive reconstruction is to noise and symmetry breaking, and whether we can construct a scalable synthetic dataset that covers a wide range of interfacial dynamics without requiring prohibitively expensive 3D simulations for every configuration. Answering these questions requires a controlled benchmark where ground-truth 4D structure is known and we can vary the number of views, noise levels, and physical parameters systematically. In the next section, we describe our methodology for creating such a benchmark by lifting cheap 2D simulations into 3D and generating physically realistic X-ray projections. This approach decouples data scarcity from algorithm development — enabling us to explore complex reconstruction models with virtually unlimited training data, and to evaluate their performance in a measurable way.

3 Synthetic Dynamic Tomography (SDT) Dataset

To facilitate learning in this domain, we construct a synthetic dataset for sparse-view 4D tomography inspired by sparse-view radiography of dynamic two-material systems, where the evolving interface (and mixing morphology) must be inferred from very few projections. The design goals are to (i) produce abundant training pairs of 3D volumes and corresponding projections, (ii) incorporate known physics (e.g. symmetries, noise characteristics) so that the learned models remain relevant to real scenarios, and (iii) cover diverse conditions beyond a single specific phenomenon. Our solution is to leverage inexpensive 2D time-dependent simulations as a starting point, then algorithmically "lift" them to 3D and generate many 2D projections from each 3D state. This process is illustrated conceptually in Figure 1.

Source simulations (2D → time): We begin with a library of 2D hydrodynamic simulations that evolve in time. For concreteness, our current dataset uses an axis-symmetric Rayleigh–Taylor instability – an accelerated two-material interface in which a heavier fluid penetrates a lighter fluid – as the example phenomenon (6; 16). Each 2D simulation produces a time series of density fields

$S_t(x, z)$ (one slice per time t) given some initial conditions. We systematically vary key dimensionless parameters of the simulation – in this case, the gravitational acceleration g, the Atwood number $\mathcal{A}$ (density contrast), and the initial perturbation amplitude a_0. These three parameters significantly affect the growth rate and morphology of the Rayleigh–Taylor instability. By simulating a grid of $(g, \mathcal{A}, a_0)$ combinations, we obtain a wide range of evolution behaviors in 2D. We emphasize that using 2D simulations is a pragmatic choice: high-fidelity 3D simulations are extremely costly, but many interfacial instabilities (especially in early-to-mid stages) exhibit approximate symmetry about an axis, meaning a 2D cross-section can capture the salient interface morphology. This 2D $\rightarrow$ 3D symmetry assumption lets us generate orders of magnitude more training samples than if we had run full 3D simulations for each set of parameters.

Symmetry-based lifting (2D $\rightarrow$ 3D): For each 2D frame S_t, we generate a plausible 3D *density/attenuation* volume V_t by revolving the 2D density field around its symmetry axis. In effect, we assume the 2D slice is a radial cross-section of the full 3D volume. This yields a dense 3D scalar field $V_t(x, y, z)$ that is consistent with the original 2D image in the axial plane. We treat this as the "ground truth" 3D state at time t. Of course, a purely symmetric revolution will produce a perfectly axisymmetric 3D volume – which may become unrealistic at later times when symmetry breaks (due to secondary instabilities or asymmetric influences). To increase diversity, we therefore create two families of 3D volumes:

1. *Symmetric volumes*: the straightforward revolution of S_t, representing an ideal case where the dynamics remain symmetric about the axis.
2. *Symmetry-broken volumes*: variants where we intentionally perturb the 3D density after lifting. For example, we might embed a small high-density sphere (a proxy for a localized material inhomogeneity) at a random location within the volume or apply a slight random deformation, introducing a controlled asymmetry. These perturbations are designed to be subtle (so that the projection images still resemble the original 2D slice to some extent) but challenging enough to test a model's robustness. They prevent a learned model from simply assuming perfect symmetry, and mimic real-world effects like experimental artifacts or slight misalignments.

By generating multiple 3D perturbation variants for each 2D frame, we effectively obtain many plausible 3D realizations of the same underlying 2D dynamics. All these volumes share the same known initial parameters $(g, \mathcal{A}, a_0)$ and time index t as the originating simulation, enriching the dataset with a variety of outcomes for a given physical configuration.

Forward projection (3D $\rightarrow$ 2D radiographs): Next, we produce synthetic X-ray projections from each 3D volume. We simulate radiography by casting rays through V_t and integrating a scalar attenuation field along each ray (here taken proportional to the simulated density). The result is a 2D image $p_\theta(u, v) = \int \mu(x, y, z)\, dl$ from view angle θ, analogous to a radiograph or computed tomography (CT) projection. For each volume, we render projections from 72 uniformly spaced view angles covering 360°. This dense set of viewpoints allows us to subsequently simulate sparse-view scenarios by selecting only k out of the 72 projections as input. (For instance, $k = 8, 4, 2$ views would correspond to an extremely sparse CT with 45°, 90°, or 180° separations between angles.) Full specifics are given in Appendix A.

4 Experiments and Results

Experimental Setup: We evaluate on the Rayleigh–Taylor dataset (section 3), containing 54 dynamic 3D sequences with 72 uniformly spaced views, divided into (1) symmetric volumes from solid-of-revolution and (2) symmetry-broken volumes with embedded high-density spheres (a proxy for localized material inhomogeneity). Models are trained on a single NVIDIA Tesla V100 GPU under sparse-view regimes and evaluated on held-out angles using PSNR and qualitative inspection (training details in Appendix B).

Symmetric Volumes: Training for 20,000 iterations (˜5 h) yields reconstructions with PSNR ≈ 35 dB on unseen views. Even from a single projection, the model recovers the spike–bubble interface and mushroom-cap morphology, maintaining temporal continuity without strong angular artifacts. Angular consistency is shown in Appendix B–Table 1, where views near 0° and 360° remain sharp while reconstructions degrade around 180°–270° due to sparse-view ambiguity.

Symmetry-Broken Volumes: Perturbed volumes require 100,500 iterations (˜30 h) to converge at PSNR $\approx$ 35 dB. Reconstructions capture the global interfacial instability and its evolution, though local fidelity near the perturbation degrades, especially in late-time single-view cases. The embedded sphere is identifiable early but partially obscured as morphological complexity increases. We further evaluate single-view refinement: as shown in Appendix B–Table 2, test-time fine-tuning steadily improves PSNR/SSIM and visual sharpness, demonstrating that latent-code adaptation can recover finer structural details.

ACKNOWLEDGMENTS

This research was supported by the Pazy Foundation. Computational support was provided by the NegevHPC project.

REFERENCES

[1] Riccardo Barbano, Chen Zhang, Simon Arridge, and Bangti Jin. Quantifying model uncertainty in inverse problems via bayesian deep gradient descent. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 1392–1399. IEEE, 2021.

[2] Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, and Angtian Wang. Structure-aware sparse-view x-ray 3d reconstruction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11174–11183, 2024.

[3] Hu Chen, Yi Zhang, Yang Chen, Jiliu Zhang, Weihua Zhang, Huaiqiang Sun, Yining Lv, Peixi Liao, Jianyu Zhou, and Ge Wang. Learn: Learned experts' assessment-based reconstruction network for sparse-data ct. *IEEE Transactions on Medical Imaging*, 37(6):1333–1347, 2018.

[4] Abril Corona-Figueroa, Jonathan Frawley, Sam Bond-Taylor, Sarath Bethapudi, Hubert PH Shum, and Chris G Willcocks. Mednerf: Medical neural radiance fields for reconstructing 3d-aware ct-projections from a single x-ray. In *2022 44th annual international conference of the IEEE engineering in medicine & Biology society (EMBC)*, pages 3843–3848. IEEE, 2022.

[5] Ivan Grega, William F Whitney, and Vikram S Deshpande. Neural rendering enables dynamic tomography. *arXiv preprint arXiv:2410.20558*, 2024.

[6] Re'em Harel, Matan Rusanovsky, Yehonatan Fridman, Assaf Shimony, and Gal Oren. Complete deep computer-vision methodology for investigating hydrodynamic instabilities. In *International Conference on High Performance Computing*, pages 61–80. Springer, 2020.

[7] Marjolein Heyndrickx, Tom Bultreys, Wannes Goethals, Luc Van Hoorebeke, and Matthieu N Boone. Improving image quality in fast, time-resolved micro-ct by weighted back projection. *Scientific Reports*, 10(1):18029, 2020.

[8] CM Huntington, KS Raman, SR Nagel, SA MacLaren, T Baumann, JD Bender, ST Prisbrey, L Simmons, P Wang, and Y Zhou. Split radiographic tracer technique to measure the full width of a high energy density mixing layer. *High Energy Density Physics*, 35:100733, 2020.

[9] Berk Iskender, Marc L Klasky, and Yoram Bresler. Red-psm: Regularization by denoising of factorized low rank models for dynamic imaging. *IEEE Transactions on Computational Imaging*, 10:832–847, 2024.

[10] Yongho Kim and Hans W Herrmann. Gamma-ray measurements for inertial confinement fusion applications. *Review of Scientific Instruments*, 94(4), 2023.

[11] Pawel Marek Kozlowski, Y Kim, Brian Michael Haines, HF Robey, Thomas Joseph Murphy, Heather Marie Johns, and Theodore Sonne Perry. Use of computer vision for analysis of image datasets from high temperature plasma experiments. *Review of Scientific Instruments*, 92(3), 2021.

[12] David S Montgomery. Invited article: X-ray phase contrast imaging in inertial confinement fusion and high energy density research. *Review of Scientific Instruments*, 94(2), 2023.

[13] Stéphane Roux, François Hild, Philippe Viot, and Dominique Bernard. Three-dimensional image correlation from x-ray computed tomography of solid foam. *Composites Part A: Applied science and manufacturing*, 39(8):1253–1265, 2008.

[14] Joshua Paul Sauppe, Yingchao Lu, Petros Tzeferacos, Adam Canady Reyes, Sasikumar Palaniyappan, Kirk Adler Flippo, Shengtai Li, and John L Kline. On the importance of three-dimensional modeling for high-energy-density physics experiments. *Physics of Plasmas*, 30(6), 2023.

[15] Katja Schwarz, Yiyi Liao, Michael Niemeyer, and Andreas Geiger. Graf: Generative radiance fields for 3d-aware image synthesis. *Advances in neural information processing systems*, 33:20154–20166, 2020.

[16] Shai Stamler-Grossman, Nadav Schneider, Gershon Hanoch, and Gal Oren. Clariphy: Physics-informed image deblurring with transformers for hydrodynamic instability analysis. In *Machine Learning and the Physical Sciences, NeurIPS*, 2024.

[17] Ge Wang, Jong Chul Ye, and Bruno De Man. Deep learning for tomographic image reconstruction. *Nature machine intelligence*, 2(12):737–748, 2020.

[18] Bradley T Wolfe, Zhizhong Han, Jonathan S Ben-Benjamin, John L Kline, David S Montgomery, Elizabeth C Merritt, Paul A Keiter, Eric Loomis, Brian M Patterson, Lindsey Kuettner, et al. Neural network for 3d icf shell reconstruction from single radiographs. *arXiv preprint arXiv:2101.03646*, 2021.

[19] Tobias Würfl, Florin C Ghesu, Vincent Christlein, and Andreas Maier. Deep learning computed tomography. In *International conference on medical image computing and computer-assisted intervention*, pages 432–440. Springer, 2016.

[20] Ruyi Zha, Yanhao Zhang, and Hongdong Li. Naf: neural attenuation fields for sparse-view cbct reconstruction. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 442–452. Springer, 2022.

[21] Chonghang Zhao, Mingyuan Ge, Xiaogang Yang, Yong S Chu, and Hanfei Yan. Limited-angle x-ray nano-tomography with machine-learning enabled iterative reconstruction engine. *npj Computational Materials*, 11(1):240, 2025.

[22] Yanjie Zheng and Kelsey B Hatzell. Ultrasparse view x-ray computed tomography for 4d imaging. *ACS Applied Materials & Interfaces*, 15(29):35024–35033, 2023.

[23] Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

A SDT Dataset Specifics

The current release of our SDT dataset consists of:

- 54 base dynamic sequences of 3D volumes, each derived from a unique combination of $(g, \mathcal{A}, a_0)$ and containing multiple time steps t. Early time steps are fairly simple (linear regime of the instability) while later steps become highly nonlinear with complex structures.
- 72 radiographs per 3D volume (one per evenly spaced angle around the object). These can be sampled to simulate any desired sparse-view configuration (e.g. using only 8 specific angles for training a model, and possibly using other angles for testing).
- Two object families: (i) perfectly symmetric volumes, and (ii) perturbed "symmetry-broken" volumes (such as those with an embedded sphere or density noise). The latter family provides a harder testbed for reconstruction algorithms, to ensure they don't rely on trivial symmetric cues.
- Rich metadata: For each sample we store the initial simulation parameters $(g, \mathcal{A}, a_0)$, the time index t, the view angle θ for each projection, and the noise level/perturbation type used. We also maintain a link to the original 2D slice S_t, although the model does not directly use S_t – it's just for reference.

- Multiple evaluation splits: We define several ways to split the data for training and testing. One split holds out entire object sequences (i.e. certain $(g, \mathcal{A}, a_0)$ combinations) to test generalization to new physical parameters. Another split holds out later time steps, so that models must extrapolate the dynamics beyond the training regime (train on early t, test on later t). We also consider a view sparsity split: for example, train a model with 8 views per time step, but evaluate it with only 4 or 2 views to see how graceful the degradation is. These varied splits encourage rigorous testing of both interpolation and extrapolation capabilities in space, time, and imaging conditions.

Overall, this dataset provides a controlled yet physically grounded benchmark for dynamic sparse-view tomography. It offers essentially unlimited training data (since we can generate more 2D simulations and repeat the lifting process arbitrarily) while avoiding the need for any real 3D ground truth from experiments. By construction, it captures the key difficulties of the real problem: very few views, temporal evolution, and the possibility of symmetry breaking and noise. Next, we describe how we leverage this dataset to train a neural field model that reconstructs the 3D+time structure.

B Radiance Field Model for 4D Reconstruction

We build our reconstruction approach on the framework of Neural Radiance Fields (NeRF), which has proven effective for recovering 3D scenes from sparse 2D images. In our context, we adopt an existing NeRF variant tailored to medical imaging called MedNeRF (4) and extend it to handle temporal sequences. In a standard NeRF, a neural network is trained to represent a static 3D volume by mapping spatial coordinates (x, y, z) (and a viewing direction) to a density and color, such that rendering this volume reproduces a set of input images. MedNeRF, in particular, incorporates learned latent codes to account for different objects' shape and appearance (disentangling those factors) and has been shown to reconstruct 3D CT volumes from as little as a single X-ray image. We modify this framework in two important ways:

- *Temporal conditioning*: Instead of modeling a static scene, we introduce time t as an additional input coordinate to the neural field. The model then represents a continuous 4D function $F_\Theta(x, y, z, t)$ that outputs the density at any spatial point and time. In practice, we implement this by feeding the time (or a learned time embedding) into the network alongside spatial coordinates. The network is trained simultaneously on projections from all time steps, effectively learning an interpolation of the volume's evolution over time. This approach allows the model to share information across time – for instance, early and late states of the same object are related, and the network can use the underlying physical continuity to better inform reconstructions at each frame. Conceptually, the model learns a dynamical radiance field: a representation that can be "queried" to obtain the 3D state at any time, which can then be projected from any angle.
- *Few-view training regime*: We train the model using only a very sparse set of projections for each time step (mirroring the eventual deployment scenario). For example, we might give the model 4 views (out of the 72 available) at each time t during training. The model is optimized to reproduce those views when rendering its internal 4D representation. We use a combination of mean squared error and structural similarity (SSIM) loss on the rendered vs. ground-truth projection images. Even though ground-truth volumes V_t exist in our dataset, we do not supervise the network with direct volume data; we treat the task exactly as it would be in reality – learning from images alone. The abundance of training samples (many different object trajectories) prevents overfitting, and the network must build a physically plausible internal model to explain each set of few projections. Notably, recent research has shown NeRF-based models can outperform traditional tomographic reconstruction from limited views, by effectively integrating information over many training examples of how objects tend to look from different angles.

Our NeRF model also incorporates latent codes similar to Generative Radiance Fields (GRAF) (15). That is, each distinct object (e.g. each simulation instance) is associated with a learned latent vector that modulates the network to account for that object's unique features. In training, these latents are optimized alongside network weights. This helps the model handle the variability across different initial conditions $(g, \mathcal{A}, a_0)$ in the dataset – rather than trying to represent vastly different evolutions with a single set of weights, the model can store object-specific information in the latent space. We reserve separate latents for shape and for appearance as in MedNeRF, although in our case "appear-

ance" mainly accounts for any per-sequence idiosyncrasies in intensity (since all our renderings are essentially density projections, color is uniform).

Single-view fine-tuning: MedNeRF suggests refine the model for a specific new scene at inference time by optimizing its latent codes to better fit the observed images, we freeze the main network weights and allow the latent vectors for shape and appearance to adjust to the new object's projections. Even a short period of this test-time finetuning (a few hundred iterations) can significantly sharpen the reconstructed images and volumes. Finetuning leads to higher PSNR (Peak Signal-to-Noise Ratio) on the predicted projections – indicating crisper, more accurate image details – and higher SSIM (Structural Similarity Index) – indicating better structural fidelity in the reconstruction. Qualitatively, this corresponds to the model nailing down the specific features of that object's shape. In other words, the pre-trained model provides a good initial guess (based on learning general physics from the training set), and the per-instance optimization then recovers the fine details that distinguish this particular object's volume. This two-stage process (learning generic features from many objects, then specializing to one object with minimal data) is crucial for robust performance in practice.

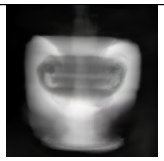
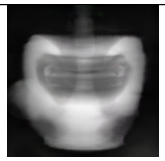
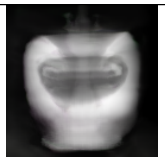
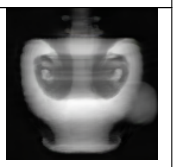
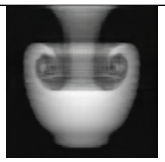
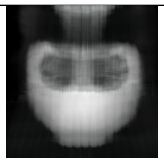
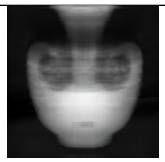
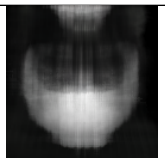
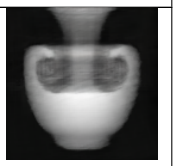
Angle	0°	90°	180°	270°	360°
Inference					
Symmetric					

Table 1: **Single-view reconstruction of an object and its symmetric counterpart across different rotation angles (0°–360°).** Each row compares the same object family, with reconstructions rendered at canonical orientations. The results illustrate how angular position affects reconstruction fidelity: views near 0° and 360° remain sharp and well structured, while intermediate rotations around 180°–270° exhibit noticeable degradation in both sharpness and appearance. This table highlights the intrinsic challenges of sparse-view tomography, where limited angular coverage leads to view-dependent artifacts.

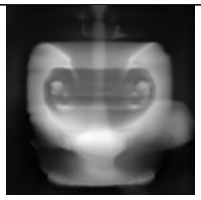
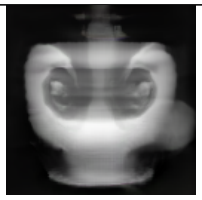
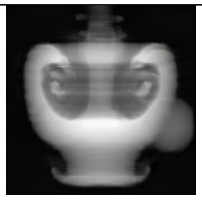
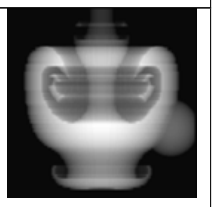
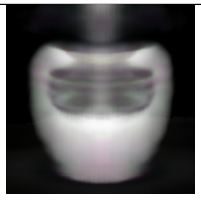
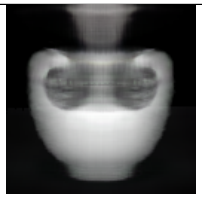
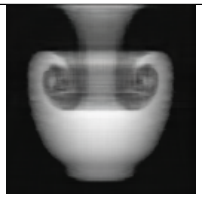
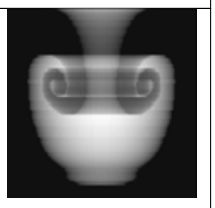
PSNR-SSIM	20-6	25-7	30-9	GT
Inference				
Symmetric				

Table 2: **Effect of single-view fine-tuning on reconstruction quality for an object and its symmetric counterpart.** Columns correspond to progressively longer optimization (with corresponding PSNR–SSIM scores) and a ground-truth (GT) reference. Fine-tuning systematically sharpens boundaries and improves morphology, with both PSNR and SSIM increasing as optimization continues. The progression demonstrates the utility of latent-code adaptation: even when starting from a weak single-view reconstruction, per-instance optimization recovers sharper details and more faithful shapes, reducing the gap to ground truth.

Observations and limitations: Our experiments reveal that sparse-view training generalizes well to novel view angles in the symmetric regime, demonstrating that the learned 4D representation can reconstruct the global spike–bubble interface morphology from limited input data. The method shows resilience to moderate symmetry breaking, maintaining coherent large-scale flow structures despite localized perturbations. However, reconstructions from a single projection, while still plausible, exhibit reduced sharpness and increased susceptibility to local errors, particularly in late-stage, highly nonlinear regimes, indicating that incorporating additional priors or physics-based constraints could improve fine-detail recovery and robustness.