Interlaced dynamic XCT reconstruction with spatio-temporal implicit neural representations

Mathias Boulanger[1, 2] and Ericmoore Jossou[1, 3]

[1]Nuclear Science and Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, Massachusetts, United States
[2]École Polytechnique de Bruxelles, Université Libre de Bruxelles, 50 Franklin Roosevelt Ave, 1050, Bruxelles, Belgium
[3]Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, 02139, Massachusetts, United States

October 13, 2025

Abstract

In this work, we investigate the use of spatio-temporal Implicit Neural Representations (INRs) for dynamic X-ray computed tomography (XCT) reconstruction under interlaced acquisition schemes. The proposed approach combines ADMM-based optimization with INCODE, a conditioning framework incorporating prior knowledge, to enable efficient convergence. We evaluate our method under diverse acquisition scenarios, varying the severity of global undersampling, spatial complexity (quantified via spatial information), and noise levels. Across all settings, our model achieves strong performance and outperforms *Time-Interlaced Model-Based Iterative Reconstruction* (TIMBIR), a state-of-the-art model-based iterative method. In particular, we show that the inductive bias of the INR provides good robustness to moderate noise levels, and that introducing explicit noise modeling through a weighted least squares data fidelity term significantly improves performance in more challenging regimes. The final part of this work explores extensions toward a practical reconstruction framework. We demonstrate the modularity of our approach by explicitly modeling detector non-idealities, incorporating ring artifact correction directly within the reconstruction process. Additionally, we present a proof-of-concept 4D volumetric reconstruction by jointly optimizing over batched axial slices, an approach which opens up the possibilities for massive parallelization, a critical feature for processing large-scale datasets.

1. Introduction

The tomographic observation of in situ dynamic phenomena such as melting and solidification of alloys is inherently ill-posed, as each sinogram encodes information from different temporal states of the evolving object. Traditional reconstruction techniques typically fail to recover high-quality time-resolved volumes under such conditions. At best, they yield temporally blurred reconstructions; at worst—when the object changes too rapidly—reconstruction becomes entirely unreliable, making it difficult to observe the microstructural changes. In this work, we explore the use of *implicit neural representations* (INRs) as a flexible framework to address this challenge. INRs represent continuous signals via neural networks, enabling images to be reconstructed directly from spatio-temporal coordinates [1]. This formulation offers several key advantages: resolution independence, low data storage memory footprint, inherent smoothness, and the ability to generalize beyond discrete grid structures [1]. These properties make INRs particularly attractive for tackling challenging inverse problems such as dynamic X-ray computed tomography (XCT) reconstruction from interlaced measurements.

Early work addressing dynamic tomography under severe sampling constraints proposed interlaced acquisition schemes that distribute projection angles across time, enabling the reconstruction of time-resolved volumes from a limited number of views per frame. For instance, Mohan et al. [2] introduced *Time Interlaced Model-Based Iterative Reconstruction* (TIMBIR), a framework that combines interlaced angular sampling with 4D *Model-Based Iterative Reconstruction* (MBIR), explicitly modeling the sensor noise statistics and detector non-idealities such as zingers and ring ar-

tifacts. By redistributing projection angles more uniformly in time, TIMBIR significantly increases temporal resolution without compromising spatial quality, outperforming traditional analytic reconstructions [2].

A complementary strategy was introduced by Zang et al. [3], who proposed a space-time tomographic framework capable of jointly reconstructing volumetric sequences and estimating dense deformation fields from X-ray projections acquired under continuous object motion. Using an interlaced low-discrepancy angular sampling scheme together with a multi-scale alternating optimization, their method aggregates information across time while compensating for inter-frame deformation, enabling high-quality reconstructions for slowly and smoothly varying shapes. Building on this work, Zang et al. [4] extended the formulation to handle high-speed, non-periodic deformations through a warp-and-project reconstruction model. This approach replaces the discretized time axis with a continuous one, assigns each projection a precise capture time, and enforces projection–volume consistency via forward and backward projection of the keyframes. The method, which also employs interlaced acquisition, adaptively inserts additional keyframes in periods of rapid motion, decouples acquisition and reconstruction frame rates, and significantly improves temporal fidelity and spatial accuracy compared to the original space-time tomography.

1.1. Implicit Neural Representations for XCT Reconstruction

Early neural implicit approaches such as CoIL [5] learn a continuous mapping from measurement coordinates to sinogram values, enabling dense and high-fidelity projection fields to be synthesized from sparse or irregular acquisitions, which can then be processed with standard reconstruction algorithms. IntraTomo [6] later pioneered the direct use of INRs for tomography by parameterizing the attenuation field as a continuous coordinate-based multilayer perceptron trained directly from projection data via differentiable ray sampling. This formulation enables high-resolution reconstructions without voxel discretization, offering memory efficiency and inherent support for arbitrary sampling patterns. NeuralCT [7] extended the direct INR approach by representing the volume as a continuous signed distance function and jointly reconstructing geometry and motion through differentiable projection operators combined with spatiotemporal regularization, enabling high-resolution, motion-compensated reconstructions from standard CT sinograms.

More recently, Neural Attenuation Fields [8] adapted INR-based tomography to sparse-view CBCT via self-supervised optimization of a continuous attenuation field using a multi-resolution hash encoding [9], achieving high-fidelity reconstructions with reduced training times compared to earlier INR methods. A similar approach has been extended to dynamic or 4D CT settings. Notably, the work of Reed et al. [10] introduces a 4D reconstruction pipeline that couples INRs with parametric motion fields to enable spatiotemporally continuous reconstructions from extremely limited angular coverage, using a differentiable Radon transform in a training-free, self-supervised framework. This method illustrates the potential of INRs not only for high-quality static reconstruction but also for handling complex temporal dynamics and motion. In a different line of work, non-periodic dynamic CT, especially for correcting non-periodic, rapid motion such as high-heart-rate cardiac imaging, has been addressed by BIRD [11], which employs a backward-warping implicit neural representation combined with diffeomorphic regularization to mitigate motion artifacts.

These methods collectively demonstrate the flexibility of INRs in XCT reconstruction tasks, ranging from sparse static recovery to dynamic and limited-angle scenarios. However, in dynamic settings, existing approaches typically assume that motion can be represented as a continuous, topology-preserving transformation, often parameterized via a diffeomorphic motion field or implemented through backward-warping. Such formulations inherently prevent modeling non-topological changes, such as the merging or splitting of structures, which traditional deformation models cannot capture. In contrast, our work aims to relax this assumption while leveraging the intrinsic strengths of interlaced acquisition scheme.

2. Problem Formulation

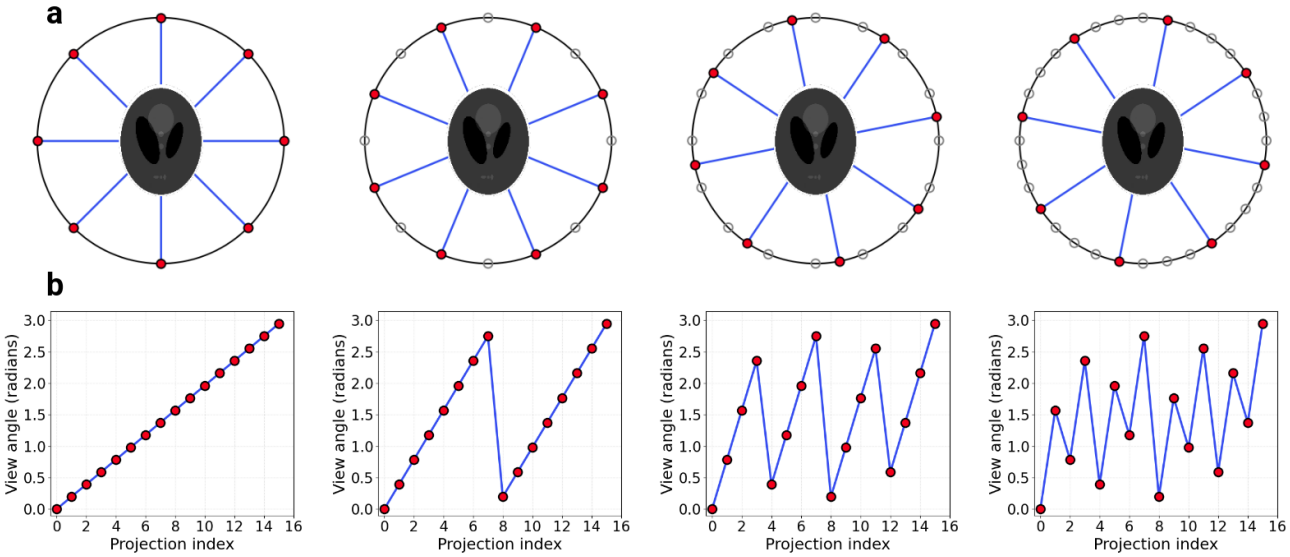
2.1. Interlaced acquisition

The conventional XCT acquisition strategy is progressive view sampling, in which projections are acquired continuously as the sample rotates at constant speed. A full set of angular views is then grouped and used to reconstruct a single static volume. In this configuration, the temporal resolution is dictated by the number of projections N_θ required for accurate spatial reconstruction. If the system acquires projections at a frequency F_c, the achievable volume frame rate becomes:

$$F_s = \frac{F_c}{N_\theta}$$

This linear relationship highlights a key bottleneck: increasing temporal resolution by reducing N_θ reduces the spatial resolution, while increasing N_θ reduces the frame rate. To mitigate this limitation, the interlaced acquisition strategy has been introduced [2], enabling improved temporal coverage without increasing the total number of projections.

In interlaced acquisition, the angular views are no longer grouped sequentially for each time frame. Instead, the full set of projection angles is distributed across multiple temporal frames in a staggered pattern. Each frame thus receives a sparse, non-contiguous subset of angular views, carefully selected to maximize angular diversity over time. This approach ensures that projections are spread as evenly as possible across both angle and time, which helps reduce aliasing and



▼ **Figure 1**: (a) Visual representation of interlaced acquisition for K = 4 and $N_\theta = 16$. In practice, it is the sample that rotates. (b) Illustration of interlaced view sampling for different values of K and $N_\theta = 16$. From left to right: K = 1, K = 2, K = 4 and K = 16.

improve temporal sampling coverage. Over multiple frames, all angular views are still required to globally agree with the spatial Nyquist theorem, but each individual frame is undersampled from a spatial perspective. A typical interlacing pattern uses a bit-reversal sequence to determine the view assignment. For interlaced sampling over π radians, the angle θ_n for the projection index n is defined as [2]:

$$\theta_n = \left[\left(n \bmod \frac{N_\theta}{T}\right) T + \mathrm{Br}\left(\left\lfloor \frac{nT}{N_\theta} \right\rfloor \bmod T\right)\right] \cdot \frac{\pi}{N_\theta}, \tag{1}$$

where N_θ is the total number of distinct projection angles, T is the number of sub-frames, and $\mathrm{Br}(\cdot)$ is the bit-reversal over $\log_2 K$ bits.

The goal of this scheme is to increase the acquisition frequency to $F_s = T\frac{F_c}{N_\theta}$. Figure 1 illustrates the principle of interlaced acquisition. The benefit of interlacing lies in its ability to decouple temporal resolution from the total number of projections, without increasing acquisition speed. However, this gain comes at the cost of incomplete angular coverage per frame, which poses significant challenges for traditional analytic reconstruction methods.

2.2. Static XCT scenario

In a static XCT scenario, suppose we are given sinogram measurements $\mathbf{y} \in \mathbb{R}^{N_\theta \times N_d}$, where N_θ denotes the number of projection angles and N_d the number of detector elements. Let $f_\theta : \mathbb{R}^n \to \mathbb{R}$ be an implicit neural representation (INR) parameterized by $\theta \in \mathbb{R}^p$, with p, the number of parameters. Let $\mathbf{x} = \mathcal{T}(f_\theta(\mathcal{G}))$ denote the image obtained by evaluating f_θ over a spatial grid $\mathcal{G} \subset \mathbb{R}^n$, discretized as a set of $H \times W$ coordinates. The goal is to learn the parameters θ such that the forward projection of f_θ matches the observed sinogram, i.e., $P\{\mathbf{x}\} \approx \mathbf{y}$, where P denotes the forward Radon transform, which computes discrete approximations to the line integrals of $\mathbf{x}$ along the ray paths defined by the scanner's acquisition geometry. This translates into the following optimization problem:

$$\min_\theta \mathcal{L}(P\{\mathbf{x}\}, \mathbf{y}), \tag{2}$$

where $\mathcal{L}$ is a differentiable loss function, typically a mean squared error.

2.3. Dynamic XCT Scenario

In the dynamic XCT scenario, we consider a sequence of sinogram measurements denoted by $\mathbf{Y} = [\mathbf{y}_0, \ldots, \mathbf{y}_{T-1}]$, where T is the number of temporal acquisitions. Each acquisition corresponds to a full gantry rotation, performed using an interlaced sparse sampling scheme.

We define the angular sampling strategy more precisely. Let $\Theta = \{\theta_1, \ldots, \theta_{N_\theta}\}$ denote the full set of projection angles. For each time step $t \in \{0, \ldots, T-1\}$, we acquire a subset $\Theta_t \subset \Theta$, with $|\Theta_t| \ll |\Theta|$, such that:

$$\bigcup_{t=0}^{T-1} \Theta_t = \Theta, \quad \text{and} \quad \Theta_t \cap \Theta_{t'} \approx \emptyset \text{ for } t \neq t'. \tag{3}$$

At each time step t, we model the dynamic volume using a time-dependent implicit neural representation (INR), $f_\theta(\cdot, t) : \mathbb{R}^n \to \mathbb{R}$. The reconstructed image is obtained by evaluating this representation over a spatial grid $\mathcal{G} \subset \mathbb{R}^n$:

$$\mathbf{x}_t = \mathcal{T}(f_\theta(\mathcal{G}, t)). \tag{4}$$

Let $\mathbf{X} = [\mathbf{x}_0, \ldots, \mathbf{x}_{T-1}]$ be the reconstructed sequence, and $\mathbf{\Theta} = \{\Theta_0, \ldots, \Theta_{T-1}\}$ the corresponding angular subsets.

The reconstruction objective is formulated as the following optimization problem:

$$\min_\theta \mathcal{L}\left(P_{\mathbf{\Theta}}\{\mathbf{X}\}, \mathbf{Y}\right), \tag{5}$$

where $P_{\boldsymbol{\Theta}}$ denotes the time-wise application of the Radon transform over the sequence of images, restricted to the respective angular subsets.

2.4. Spatio-temporal regularization

To promote both spatial and temporal smoothness in the reconstructed sequence, we incorporate total variation (TV) regularization terms in both space and time domains [12].

The spatial TV encourages piecewise-smoothness within each frame, independently across time:

$$\mathrm{TV}_{\mathrm{space}}(\mathbf{X}) = \frac{1}{T}\sum_{t=0}^{T-1} \|\nabla \mathbf{x}_t\|_1 \,. \tag{6}$$

The temporal TV penalizes abrupt variations over time by promoting smooth transitions between consecutive frames:

$$\mathrm{TV}_{\mathrm{time}}(\mathbf{X}) = \frac{1}{T-1}\sum_{t=0}^{T-2} \|\mathbf{x}_{t+1} - \mathbf{x}_t\|_1 \,. \tag{7}$$

The final optimization problem is formulated as:

$$\begin{aligned}\min_{\theta}\ \mathcal{L}\left(P_{\boldsymbol{\Theta}}\{\mathbf{X}\}, \mathbf{Y}\right) + \lambda_s\, \mathrm{TV}_{\mathrm{space}}(\mathbf{X}) \\ + \lambda_t\, \mathrm{TV}_{\mathrm{time}}(\mathbf{X}),\end{aligned} \tag{8}$$

where λ_s and λ_t are regularization weights controlling the strength of spatial and temporal smoothing, respectively.

3. Implementation

3.1. Encoding

A central limitation of implicit neural representations (INRs) is their spectral bias, the tendency to favor low-frequency functions, which impairs their ability to capture high-frequency details [1, 13]. A common remedy is positional encoding, which maps input coordinates to higher-dimensional spaces via sinusoidal or randomized projections to expand frequency capacity [9, 14]. For instance, a basic Fourier feature like $\gamma(x) = [\cos(2\pi x), \sin(2\pi x)]^T$ introduces periodicity. More advanced encodings of the form $\gamma(x) = [x, \ldots, \cos(2\pi\omega^{j/m}x), \sin(2\pi\omega^{j/m}x)]^T$, with ω a frequency hyperparameter and m the embedding dimension, further increase expressiveness [1].

While effective, these encodings are task-agnostic and fixed, lacking adaptation to signal structure. To address this, Implicit Neural Conditioning with Prior Knowledge Embeddings (INCODE) [15] introduces a conditioning mechanism that modulates the network's frequency response based on prior knowledge.

INCODE departs from traditional encodings by integrating a latent representation, extracted via pretrained models, that controls the network behavior. It employs a dual-network design: a *composer* MLP with generalized sinusoidal activations, and a *harmonizer* that predicts activation parameters from a latent code. The resulting activation function is:

$$f(x) = \mathbf{a} \cdot \sin(\mathbf{b}\omega_0 x + \mathbf{c}) + \mathbf{d}, \tag{9}$$

where $\mathbf{a}, \mathbf{b}, \mathbf{c}, \mathbf{d}$ respectively control the amplitude, frequency, phase, and offset. These are dynamically predicted by the harmonizer from a pretrained encoder such as ResNet, enabling task-adaptive modulation of the frequency response.

In our experiments, we evaluated several encoding strategies to enhance the representative capacity of the INR, including Fourier Feature Mapping [14], Instant-NGP-style encodings [9], and learned latent code conditioning. Among these, the INCODE architecture [15] demonstrated the most promising results in terms of reconstruction quality and convergence speed, while maintaining a low parameter count in our dynamic tomography setting.

3.2. ADMM-based optimization

To accelerate convergence during training, we adopt an ADMM-based optimization strategy developed in [16], reformulated for our dynamic setting. Instead of directly minimizing the reconstruction loss over the INR parameters θ, we decouple the data fidelity term from the implicit representation using auxiliary variables. For a static scenario, the minimization scheme is as follows:

$$\min_{\theta}\ \frac{1}{2}\|\mathbf{P}\left(\mathcal{T}(f_\theta(\mathcal{G}))\right) - \mathbf{y}\|_2^2, \tag{10}$$

and is equivalently rewritten in a constrained form:

$$\min_{\theta,\mathbf{x}}\ \frac{1}{2}\|\mathbf{P}\mathbf{x} - \mathbf{y}\|_2^2 \quad \text{subject to} \quad \mathbf{x} = \mathcal{T}(f_\theta(\mathcal{G})), \tag{11}$$

which gives the following rescaled augmented Lagrangian:

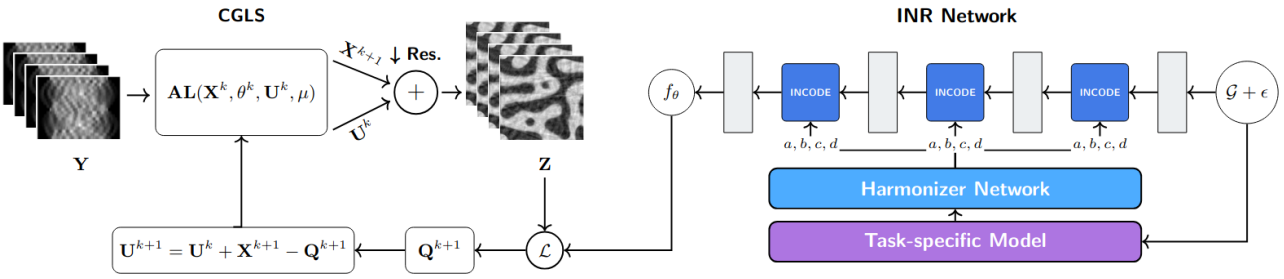
$$\begin{aligned}AL(\mathbf{x}, \theta, \mathbf{u}, \mu) =& \frac{1}{2}\|\mathbf{P}\mathbf{x} - \mathbf{y}\|_2^2 \\ &+ \frac{\mu}{2}\|\mathbf{x} - \mathcal{T}\{f_\theta(\mathcal{G})\} + \mathbf{u}\|_2^2.\end{aligned} \tag{12}$$

where $\mathbf{u} \in \mathbb{R}^n$ has the interpretation as a vector of Lagrange multipliers. We extend the original formulation to the dynamic case by solving a sequence of T such problems, one per time frame, while sharing the INR parameters across time. At each ADMM iteration k, the updates are as follows:

- **x-update (frame-wise):** for each $t = 0, \ldots, T-1$, we solve

$$\begin{aligned}\mathbf{x}_t^{k+1} = \arg\min_{\mathbf{x}}\ &\frac{1}{2}\|P_{\Theta_t}\mathbf{x} - \mathbf{y}_t\|_2^2 \\ &+ \frac{\mu}{2}\|\mathbf{x} - \mathbf{q}_t^k + \mathbf{u}_t^k\|_2^2,\end{aligned} \tag{13}$$

  which is a regularized least-squares problem solved efficiently using conjugate gradient least square (CGLS).



▼ **Figure 2**: Network architecture. The reconstruction pipeline begins with an ADMM pass, where the input sinograms are used to compute an updated sequence of image estimates $\mathbf{X}$. These estimates are then combined with the Lagrange multipliers to form the intermediate variable $\mathbf{Z}$, which is subsequently downsampled. The INR optimization phase then begins. A coarse evaluation grid $\mathcal{G}$, perturbed with random offsets ϵ, is used to sample the coordinates and evaluate the neural model's output. An illustration of the INR architecture is provided, explicitly incorporating the INCODE module. Standard neural network layers are shown in gray. At each evaluation step, the output image generated by the INR is compared to the downsampled version of $\mathbf{Z}$. The loss is computed from this comparison, along with spatial and temporal regularization terms. Once the model parameters are updated, the auxiliary variables $\mathbf{Q}$, which represent the INR outputs and $\mathbf{U}$ dual variables, are also updated. This iterative cycle continues until convergence.

- **θ-update (shared across time):** we update the INR parameters by minimizing the mismatch with the current x_t estimates and applying spatio-temporal regularization:

$$\theta^{k+1} = \arg\min_{\theta} \mathcal{L}(\mathbf{X}, \mathbf{Z}) + \lambda_s \mathrm{TV}_{\mathrm{space}}(\mathbf{X}) + \lambda_t \mathrm{TV}_{\mathrm{time}}(\mathbf{X}), \quad (14)$$

  here $\mathbf{X} = [\mathbf{x}_t]_{t=0}^{T-1}$ denotes the sequence of reconstructed images over time, where each frame $\mathbf{x}_t \in \mathbb{R}^n$ is generated from the INR. Similarly, $\mathbf{Z} = [\mathbf{x}_t^{k+1} + \mathbf{u}_t^k]_{t=0}^{T-1}$ combines the current estimate and dual variable for each frame, and is used as target in the parameter update step.

- **q-update:** for each $t = 0, \ldots, T-1$, we evaluate the current INR:

$$\mathbf{q}_t^{k+1} = \mathcal{T}(f_{\theta^{k+1}}(\mathcal{G}, t)). \quad (15)$$

- **u-update:** for each $t = 0, \ldots, T-1$, dual variables are updated via:

$$\mathbf{u}_t^{k+1} = \mathbf{u}_t^k + \mathbf{x}_t^{k+1} - \mathbf{q}_t^{k+1}. \quad (16)$$

This alternating scheme allows us to decouple the projection-heavy x updates from INR training, significantly accelerating convergence. The CGLS algorithm requires access to an adjoint operator P^T that is mathematically consistent with the forward projection P. However, in practice, many available implementations of the Radon forward and back-projection operators, such as those provided by *scikit-image* [17] and *torch-radon* [18], do not satisfy the mathematical adjointness condition. As a result, the fundamental relationship given by:

$$\langle P\mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{x}, P^T\mathbf{y} \rangle \quad (17)$$

may be violated numerically, compromising a core assumption that underpins the convergence guarantees of CGLS.

To address this, we implemented custom forward and adjoint operators using the `PyLops` [19] library and GPU-accelerated computations with `CuPy`, inspired by implementations commonly adopted in tomographic reconstruction literature [20, 21]. To ensure memory-efficient interoperability between PyTorch and CuPy, we leverage the `DLPack` protocol, enabling zero-copy, in-place sharing of GPU tensors across frameworks.

3.3. Improving memory efficiency during training

We applied two practical architectures to ensure memory efficiency during model training. Following the jittered sampling approach of [10, 22], we improve scalability by evaluating the INR on a lower-resolution spatial grid at each optimization step. To preserve the continuity of the implicit signal, random perturbations called jitters are applied to the coarse spatial coordinates. The CGLS iterations are still made at full resolution, allowing details preservation.

To further address memory constraints during training, we apply the backward pass separately for each frame rather than accumulating gradients across the entire sequence. As a consequence, temporal regularization is computed in a pairwise fashion between consecutive frames $(t-1, t)$, which allows us to avoid retaining all intermediate frames within the computational graph. To mitigate directional bias and to enforce a more symmetric temporal consistency, we alternate the processing order of the sequence at each outer iteration, traversing it in chronological order for even iterations and in reverse (anti-chronological) order for odd ones. This simple strategy promotes bidirectional temporal smoothness without increasing memory requirements. This improves the average performance across frames while reducing variability, particularly for the first frame, which otherwise lacks temporal regularization.

Figure 2 provides an overview of the ADMM-based optimization scheme and the INR architecture, while the pseudo-code is presented in Algorithm 1.

Algorithm 1 ADMM-INR algorithm

Input: Temporal sinograms $\mathbf{Y} = \{\mathbf{y}_t\}_{t=0}^{T-1}$
Init: θ^0 (INR), $\mathbf{x}_t^0 \leftarrow \mathbf{FBP}_{\Theta_t}(\mathbf{y_t})$, $\mathbf{u}_t^0 \leftarrow \mathbf{0}$, $\mathbf{q}_t^0 \leftarrow \mathcal{T}(f_{\theta^0}(\mathcal{G}, t))$ for $t = 0 \dots T-1$
for $t = 0$ **to** $T-1$ **do**
  $\mathbf{x}_t^{k+1} \leftarrow \arg\min_{\mathbf{x}} \frac{1}{2}\|P_{\Theta_t}\mathbf{x} - \mathbf{y}_t\|_2^2 + \frac{\mu}{2}\|\mathbf{x} - \mathbf{q}_t^k + \mathbf{u}_t^k\|_2^2$ ▷ CGLS
end for
Set frame order $\mathcal{O}_k$ ▷ chronological if k even, reverse if k odd
$\bar{\mathbf{p}}_{t-1} \leftarrow \varnothing$ ▷ previous prediction
for $t \in \mathcal{O}_k$ **do**
  $\tilde{\mathbf{z}}_t \leftarrow \mathcal{D}_s(\mathbf{x}_t^{k+1} + \mathbf{u}_t^k)$ ▷ downsampled target
  $\mathbf{p}_t \leftarrow \mathcal{T}(f_\theta(\tilde{\mathcal{G}}, t))$ ▷ INR output (low-res)
  $\mathcal{L} \leftarrow \|(\mathbf{p}_t - \tilde{\mathbf{z}}_t)\|_2^2$ ▷ MSE
  if $k > k_s$ **then**
    $\mathcal{L} \leftarrow \mathcal{L} + \lambda_s \|\nabla \mathbf{p}_t\|_1$ ▷ spatial TV
  end if
  if $\bar{\mathbf{p}}_{t-1} \neq \varnothing$ **and** $k > k_t$ **then**
    $\mathcal{L} \leftarrow \mathcal{L} + \lambda_t \|\mathbf{p}_t - \bar{\mathbf{p}}_{t-1}\|_1$ ▷ temporal TV
  end if
  backward$(\mathcal{L})$
  $\bar{\mathbf{p}}_{t-1} \leftarrow \text{detach}(\mathbf{p}_t)$
end for
for $t = 0$ **to** $T-1$ **do**
  $\mathbf{q}_t^{k+1} \leftarrow \mathcal{T}(f_{\theta^{k+1}}(\mathcal{G}, t))$
end for
for $t = 0$ **to** $T-1$ **do**
  $\mathbf{u}_t^{k+1} \leftarrow \mathbf{u}_t^k + \mathbf{x}_t^{k+1} - \mathbf{q}_t^{k+1}$
end for
return θ^K

4. Experiments

4.1. Simulated X-ray computed tomography data set

To evaluate the proposed framework under controlled yet physically plausible dynamics, we simulate a time-evolving microstructural process governed by spinodal decomposition [23]. This phenomenon describes phase separation in binary mixtures and is modeled by the Cahn–Hilliard equation [24]:

$$\frac{\partial c(\mathbf{r}, t)}{\partial t} = \nabla \cdot \left[M(c) \nabla \left(\frac{\partial f}{\partial c} - \epsilon \nabla^2 c \right) \right], \qquad (18)$$

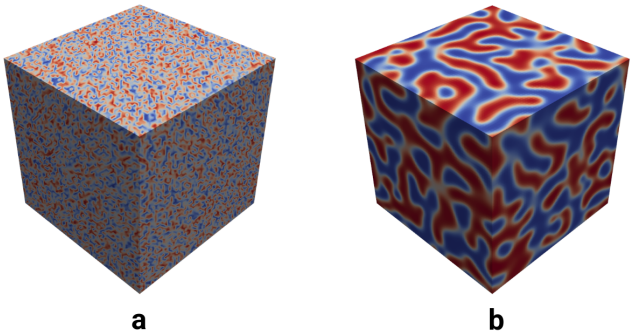
here, $c(\mathbf{r}, t)$ denotes the concentration field, $f(c)$ is the chemical free energy density, $M(c)$ the mobility, and ϵ a gradient energy coefficient controlling interface width.

Spinodal decomposition offers a rich yet controlled and reproducible spatio-temporal dynamic, characterized by evolving interfaces and microstructural complexity. It has been used in prior work to test dynamic CT frameworks [2].

Beyond serving as a synthetic benchmark, spinodal decomposition constitutes a fundamental mechanism in materials science, enabling the design of architected materials with tailored properties such as high energy absorption [25] or extreme mechanical resilience [26]. It has also been leveraged to engineer battery electrodes with enhanced capacity and long term cycling stability [27, 28]. Its influence even extends to nuclear materials, where it modulates the thermomechanical properties of metallic fuels under irradiation [29, 30].

For the 2D case, we use the open-source solver Prism-pf, which employs a matrix-free finite element method for efficient numerical computation [31]. This produces a sequence of N high-resolution frames representing the evolving object over time. Meanwhile, for 3D data generation (Figure 3), we used a Python-based implementation using a semi-implicit spectral method [32], as the use of Prism-pf proved computationally and memory intensive, making data processing impractical.

To emulate dynamic XCT acquisition with motion, the sequence of images is divided into T temporal groups of $P = N/T$ frames, each assigned a unique set of projection angles $\Theta_t = \{\theta_1^{(t)}, \dots, \theta_P^{(t)}\} \subset \Theta$ following the interlaced acquisition scheme. In each group, the i-th frame is projected at angle $\theta_i^{(t)}$, producing a sinogram composed of projections of different object states. As a representative case to assign realistic X-ray attenuation, we use the Al–Cu alloy system (matrix α-Al and Al_2Cu precipitate). The voxel resolution is $3{\times}3{\times}3$ μm^3, and at 60 keV the two phases have attenuation coefficients of 0.0750 mm^{-1} and 0.4303 mm^{-1}, respectively [33]. For visualization and evaluation, we define the ground truth at time t as the filtered back projection (FBP) of the central frame of group t, approximating the average object state.



▼ **Figure 3**: 3D Phase-field simulation of spinodal decomposition performed using a semi-implicit spectral method. (a) Initial , (b) final microstructure respectively. The simulation domain consists of a $64{\times}64{\times}64$ grid, with a constant mobility M of 1, and an initial homogeneous composition c of 0.5.

4.2. Model evaluation metrics

We evaluate reconstruction quality using two standard metrics: peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM). PSNR quantifies pixel-wise fidelity, while SSIM captures perceptual similarity.

PSNR is defined as:

$$\mathrm{PSNR}(x, x^*) = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}(x, x^*)}\right), \tag{19}$$

where x is the reconstruction, x^* the ground truth, and MAX the maximum pixel value. The mean squared error (MSE) is given by:

$$\mathrm{MSE}(x, x^*) = \frac{1}{N}\sum_{i=1}^{N}(x_i - x_i^*)^2. \tag{20}$$

SSIM [34] evaluates structural similarity over local patches:

$$\mathrm{SSIM}(x, x^*) = \frac{(2\mu_x\mu_{x^*} + C_1)(2\sigma_{xx^*} + C_2)}{(\mu_x^2 + \mu_{x^*}^2 + C_1)(\sigma_x^2 + \sigma_{x^*}^2 + C_2)}, \tag{21}$$

where μ, σ^2 and σ_{xx^*} denote local means, variances, and covariance; C_1, C_2 are small constants to stabilize the expression.

Together, PSNR and SSIM provide a complementary assessment of reconstruction quality, from both pixel-level accuracy and perceptual consistency.

4.3. Experimental setup

All experiments were performed using PyTorch on an NVIDIA GeForce RTX 4060 GPU with 8GB of memory. We employ the Adam optimizer [35] with a learning rate scheduler that decreases the learning rate progressively to support convergence. Each ADMM iteration consisted of 20 CGLS steps followed by 50 INR updates. Unless otherwise stated, the INR was implemented as a three-layer multilayer perceptron (MLP) with 256 hidden units per layer, using Gaussian positional encoding with a scale factor of 5 and a mapping input of 256. The INCODE harmonizer was truncated at the fifth order and employed the SiLU activation function. A pretrained ResNet34 from the PyTorch model zoo served as the encoder.

For TIMBIR reconstructions, the parameters of the q-generalized Gaussian Markov random field (qGGMRF) prior were tuned to optimize reconstruction quality. In our case, we adjusted the spatio-temporal regularization hyperparameters, with all configurations detailed in the accompanying code (sec. 10). As our method is iterative, we track the mean residual $\mathbf{x} - \mathbf{q}$ across temporal frames at each iteration and retain the model weights corresponding to the iteration that minimizes this residual.

5. Results and discussion

5.1. First experiment

Let δt denote the acquisition time per projection. We consider three configurations that span the same total acquisition duration $T = N_\theta \cdot \delta t$, corresponding to the full temporal evolution of the dynamic process. Specifically, we set $(N_\theta, \delta t)$ to $(256, \delta t)$, $(128, 2\delta t)$, and $(64, 4\delta t)$, respectively, with $K = 16$ fixed in all cases. By increasing the acquisition time per projection in the sparser settings, we maintain the same temporal range from the initial to the final state across configurations. This allows us to assess reconstruction robustness under decreasing angular sampling density, while keeping the dynamic acquisition window fixed. For consistency with TIMBIR, which, to our understanding, is constrained to circular reconstruction domains, we restrict the number of detectors to the image width.

Figure 4 and Table 1 reports reconstruction performance across acquisition settings with decreasing angular sampling density. As the number of projections is reduced from 256 to 64, both methods show a drop in performance, reflecting the increased difficulty of the inverse problem under sparse angular coverage. For TIMBIR, PSNR decreases from 16.57dB to 14.14dB, and SSIM from 0.723 to 0.512. In contrast, our method shows a smaller decline, from 24.83dB to 22.86dB in PSNR, and from 0.903 to 0.861 in SSIM.

Across all configurations, our approach maintains a consistent advantage of approximately 7–9dB in PSNR and 0.15–0.35 in SSIM over TIMBIR. While both methods degrade as angular sampling becomes sparser, the relative stability of our model suggests that the implicit representation is able to better exploit spatio-temporal continuity in the data. These results highlight the potential of INR-based reconstruction strategies to maintain quality even under challenging acquisition conditions.

5.2. Impact of microstructural complexity

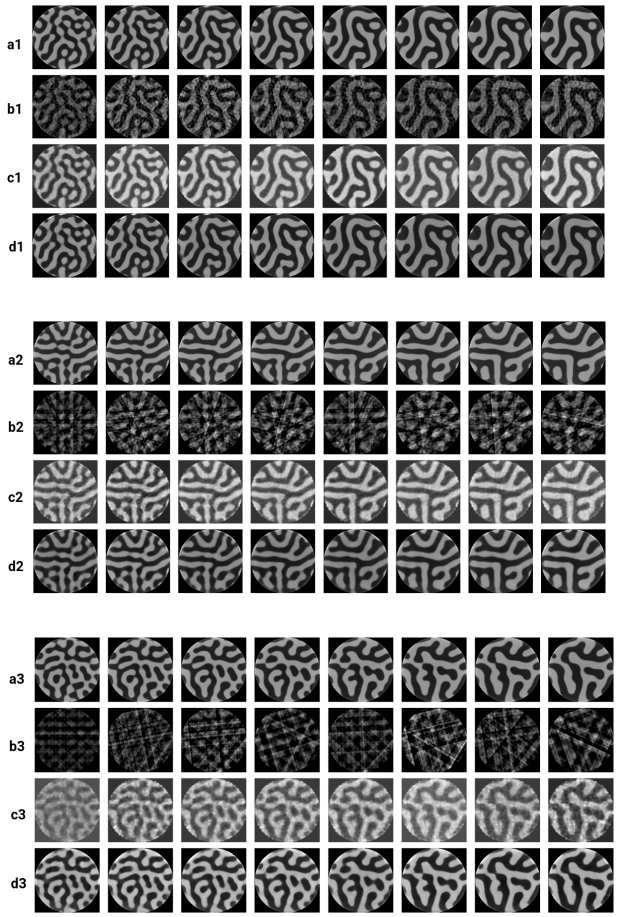
Understanding the behavior of the model under varying microstructural complexity is essential for robust performance analysis. While the previous results provided promising insights, they did not explicitly account for the intrinsic spatial complexity of the input data. In this section, we investigate how increasing microstructural complexity affects model behavior.

Previous simulations were generated on a high-resolution mesh and subsequently mapped onto images. In earlier experiments, only a cropped 256×256 region of these high-resolution images was used as input, thereby limiting the spatial variability and fine-scale details present in the data. In the present case, while the model input size remains fixed at 256×256, the cropped regions are now extracted from larger areas of 512×512 and 1024×1024, respectively. This increase in crop size enhances the structural complexity and information present in the input, allowing us to evaluate the sensitivity of the model to more complex and diverse textures and longer-range spatial dependencies.

Quantifying image complexity, however, is not a straightforward task. Classical information-theoretic measures such as Shannon entropy are ill-suited because entropy is calculated without considering spatial structures [36]. To address this, we rely on the concept of spatial information (SI) [36]. SI is based on the magnitude of local spatial gradients and captures

▲ Table 1: Reconstruction performance across acquisition configurations. Bold indicates best performance.

	$(256,\ \delta t)$		$(128,\ 2\delta t)$		$(64,\ 4\delta t)$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TIMBIR	16.57 ± 2.30	0.723 ± 0.029	16.54 ± 1.45	0.669 ± 0.027	14.14 ± 1.08	0.512 ± 0.020
Ours	**24.83 ± 0.60**	**0.903 ± 0.008**	**23.23 ± 0.60**	**0.881 ± 0.014**	**22.86 ± 0.85**	**0.861 ± 0.025**



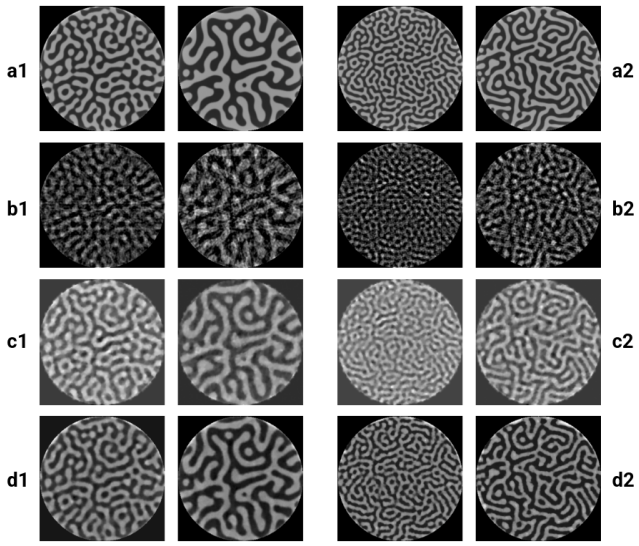
▼ **Figure 4**: Reconstruction results under varying acquisition settings. **1** | Full sampling with $N_\theta = 256$, **2** | Moderate undersampling with $N_\theta = 128$, **3** | Severe undersampling with $N_\theta = 64$. (a) Ground truth, (b) FBP reconstruction, (c) TIMBIR reconstruction, (d) Our method. Although $K = 16$ time frames were reconstructed, only every second frame is shown to reduce visual clutter.

the amount of local structural variation present in an image. Let s_h and s_v denote gray-scale images filtered with horizontal and vertical Sobel kernels, respectively. Given an image of P pixels, the mean magnitude of SI at every pixel is given by:

$$SI_r = \sqrt{s_h^2 + s_v^2}, \quad SI_{\text{mean}} = \frac{1}{P} \sum SI_r. \tag{22}$$

To ensure contrast invariance, SI is computed on a binarized version of the image. For temporal sequences, we retain the maximum SI_{mean} across all frames. For these cases we increase the Gaussian positional encoding to 10 and 15 with a mapping input size of 512.

Figure 5 and Table 2 reports reconstruction performance across two levels of spatial complexity, measured using the Spatial Information (SI) metric. When the



▼ **Figure 5**: Impact of spatial complexity on reconstruction performance. **1** | Case 1, **2** | Case 2. (a) Ground truth, (b) FBP reconstruction, (c) TIMBIR reconstruction, (d) Our method. Although $K = 16$ time intervals were reconstructed, we only show the first and last ones to reduce visual clutter.

spatial content is less complex (SI = 0.724), both methods perform better overall, but our approach shows a notable improvement over TIMBIR, with a PSNR gain exceeding 3 dB and a SSIM increase of nearly 0.06. In the more challenging setting (SI = 1.056), performance drops for both methods, as expected, due to the increased structural variability. Nevertheless, our method maintains a consistent advantage in both PSNR and SSIM.

Importantly, our approach also exhibits reduced temporal variability, as evidenced by the lower standard deviation values, particularly in the high-complexity case. This suggests that the method not only achieves higher average fidelity, but does so more consistently across the temporal sequence. Overall, these results indicate that the method remains robust across varying levels of spatial detail, while maintaining temporal stability.

5.3. Noise resilience

While the previous XCT scan model assumes ideal, noise-free measurements, real-world acquisitions are inherently affected by noise, particularly under low-dose conditions, where photon counts are limited and detector imperfections may be present. To evaluate the robustness of our model under realistic noise conditions, we simulate X-ray acquisition noise using a Poisson

▲ **Table 2**: Reconstruction performance for varying crop sizes and spatial complexity. SI = Spatial Information. Bold indicates best result per metric.

	SI = 0.724		SI = 1.056	
	PSNR	SSIM	PSNR	SSIM
TIMBIR	17.45± 1.27	0.729 ± 0.015	16.33 ± 1.47	0.730 ± 0.038
Ours	**20.78 ± 0.30**	**0.786 ± 0.009**	**17.98 ± 0.22**	**0.742 ± 0.011**

model consistent with the stochastic nature of photon detection. Given a noise-free sinogram value p, the expected photon count along a ray is modeled as $\mu = \text{dose} \cdot e^{-p}$, and noisy measurements are drawn from $I \sim \text{Poisson}(\mu)$. The corresponding noisy sinogram is then obtained by:

$$p_{\text{noisy}} = \log\left(\frac{\text{dose}}{I}\right). \tag{23}$$

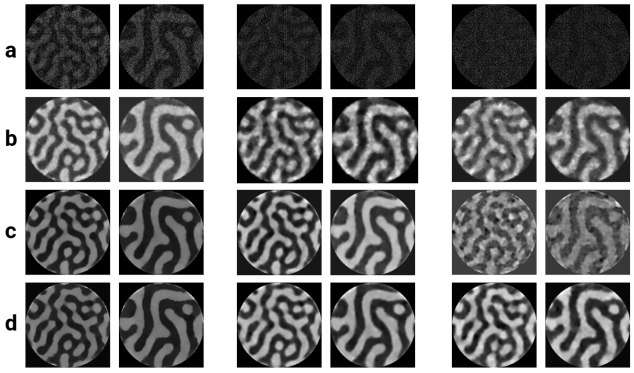
Noise severity is controlled via the dose parameter, with lower doses inducing a low signal-to-noise ratio. We evaluate three noise levels using dose values of 20×10^3, 5×10^3, and 1×10^3, applied under the first acquisition configuration (Section 5.1). This setup enables us to evaluate whether the inductive bias of INRs alone is sufficient to ensure robustness against noise. Additionally, we consider a second formulation based on a weighted least squares (WLS) model, which explicitly incorporates the statistical nature of Poisson noise. In this setting, the data fidelity term becomes:

$$\frac{1}{2}\|W^{1/2}(\mathbf{P}\mathbf{x} - \mathbf{y})\|_2^2, \tag{24}$$

where $W = \text{diag}(w)$ is a diagonal matrix with weights set proportional to the measured photon counts, i.e., $w \propto I$. This choice approximates the inverse noise variance under a Poisson model and better reflects the heteroscedastic nature of the data, thereby improving robustness in low-dose settings where the noise level varies significantly across detectors.

Figure 6 and Table 3 present the corresponding reconstruction results. As shown in Table 3, our original formulation (*Ours*) performs best at the highest dose level, achieving the highest SSIM (0.873) and PSNR (21.49 dB). This suggests that the inherent inductive bias of the INR is sufficient to ensure high-quality reconstruction in low-noise conditions.

However, as the dose decreases, noise becomes more prominent and reconstruction becomes more challenging. In these regimes, the WLS-based formulation (*Ours**), which introduces a data fidelity term weighted according to the estimated noise variance, demonstrates significantly improved robustness. At a dose of 5×10^3, *Ours** improves the SSIM from 0.778 (*Ours*) to 0.799, while also yielding the best PSNR (21.26 dB). The benefit becomes even more pronounced in the most challenging scenario of 1×10^3, where *Ours** achieves a SSIM of 0.713, substantially outperforming both the original formulation (0.559) and TIMBIR (0.556), despite TIMBIR achieving a slightly higher PSNR (17.07 dB vs. 16.81 dB). Interestingly, in these experience, TIMBIR maintain the lowest variability for PSNR results.



▼ **Figure 6**: Reconstruction of configuration 1 for different levels of dose. From left to right, dose = $\{20, 5, 1\}\times 10^3$. (a) FBP reconstruction, (b) TIMBIR reconstruction, (c) Our method, (d) Ours* = Ours with WLS. Although $K = 16$ time intervals were reconstructed, we only show the first and last reconstructions to reduce visual clutter.

6. Toward a complete framework

6.1. Detector non-idealities

Systematic detector imperfections can give rise to structured artifacts in computed tomography, most notably concentric rings in the reconstructed volume. These ring artifacts (Figure 7.a) originate from persistent inconsistencies across detector bins—such as gain variations or electronic drift, and manifest in the sinogram domain as additive biases that are invariant across projection angles but vary along the detector axis [37].

To model these effects, we introduce a vector $\mathbf{c} \in \mathbb{R}^{N_d}$ representing a static, angle-invariant additive bias for each detector bin. At each time step t, a linear operator C_t replicates $\mathbf{c}$ across the angular dimension, yielding the following forward model:

$$\mathbf{y}_t = P_{\Theta_t}\mathbf{x}_t + C_t\mathbf{c}, \tag{25}$$

where $\mathbf{y}_t$ is the measured sinogram and $P_{\Theta_t}\mathbf{x}_t$ is the ideal projection at time t.

This model is incorporated into our ADMM-based optimization by treating $\mathbf{c}$ as an auxiliary variable. At each iteration, after updating all $\mathbf{x}_t$, we compute the sinogram residuals:

$$\mathbf{r}_t = \mathbf{y}_t - P_{\Theta_t}\mathbf{x}_t^{(k+1)}, \tag{26}$$

and stack them across time and angles to form a residual matrix $R \in \mathbb{R}^{M \times N_d}$, where $M = \sum_t |\Theta_t|$ is the total number of rays. Each column of R aggregates the

▲ **Table 3**: Reconstruction performance at various dose levels (10^3). Ours* = Ours with WLS. Bold indicates best performance.

	Dose = 20		Dose = 5		Dose = 1	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
TIMBIR	17.57 ± 0.78	0.752 ± 0.012	18.39 ± 0.45	0.587 ± 0.013	**17.067± 0.612**	0.556 ± 0.022
Ours	**21.49 ± 0.86**	**0.873 ± 0.011**	19.07 ± 1.53	0.778 ± 0.021	16.68 ± 1.605	0.559 ± 0.029
Ours*	21.45 ± 1.39	0.865 ± 0.012	**21.26 ± 1.07**	**0.799 ± 0.012**	16.81 ± 1.08	**0.713 ± 0.022**

residuals corresponding to a single detector bin over all angles and time frames.

To robustly estimate $\mathbf{c}$ from R, we solve the following weighted least-squares problem using Iteratively Reweighted Least Squares (IRLS) with a Huber loss [38]:

$$\mathbf{c}^{(k+1)} = \arg\min_{\mathbf{c}} \sum_{m=1}^{M} \sum_{d=1}^{N_d} \rho_\delta(R_{m,d} - c_d), \tag{27}$$

where ρ_δ is the standard Huber function with threshold δ. The IRLS procedure assigns lower weights to outliers, improving robustness [39, 40, 41]. We enforce a zero-mean constraint on $\mathbf{c}^{(k+1)}$ to ensure identifiability and optionally apply one-dimensional smoothing (e.g., Tikhonov regularization) along the detector axis to promote spatial coherence. Alternatively, sparsity-promoting constraints could also be applied to $\mathbf{c}$ [42, 43].

Delegating artifact reduction directly to the INR, for example, through an unsupervised multi-parameter inverse solving strategy [44], would be an appealing alternative. However, due to the variable splitting introduced by the ADMM framework, the projection operator is no longer explicitly available within the INR module.

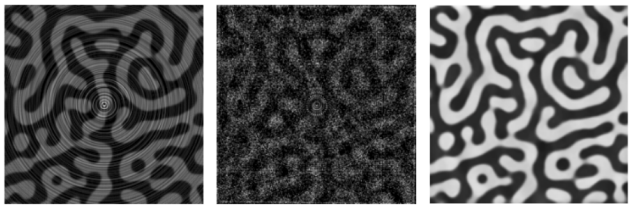
The $\mathbf{x}_t$-update step in our ADMM framework solves the following regularized least-squares problem:

$$\mathbf{x}_t^{k+1} = \arg\min_{\mathbf{x}} \frac{1}{2} \left\| P_{\Theta_t}\mathbf{x} - (\mathbf{y}_t - C_t\mathbf{c}^k) \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{x} - \mathbf{q}_t^k + \mathbf{u}_t^k \right\|_2^2. \tag{28}$$

While we retain the standard quadratic fidelity term in this work, it may be replaced by a robust alternative such as the Huber loss [38, 2] to ensure comprehensive treatment of zingers. This modification remains fully compatible with our ADMM-INR framework.

6.2. Four-dimensional extension

The previous sections and results have focused primarily on reconstructing 2D+t volumes. However, this approach can be extended to full four-dimensional (4D) volumetric reconstructions by incorporating the axial coordinate z into the INR, which now takes spatio-temporal coordinates (x, y, z, t) as input. To enforce spatial consistency along the axial direction, the loss function is augmented with an additional axial continuity term, and the optimization loop includes an iteration over z.



▼ **Figure 7**: Ring artifacts. (a) FBP reconstruction from a full-view ($N_\theta = 256$) acquisition on a static image, illustrating the strength of the injected ring artifacts. This provides a visual benchmark for detector-induced bias. (b) FBP reconstruction of the first individual subframe in a temporally interlaced setting with $K = 16$ subframes, where the reduced angular density makes artifacts less apparent. (c) Reconstruction using our method under the same acquisition setup. A detector count of $N_d = 363$ ensures complete coverage.

Nevertheless, training a single INR over the entire 3D+t volume is not computationally efficient. First, axial continuity provides meaningful regularization only within a limited spatial range. Second, optimizing a monolithic INR restricts parallelism and requires substantial computational resources on large datasets. We therefore adopt a more scalable strategy based on axial batching. Instead of reconstructing the full 4D volume at once, we divide it into smaller subproblems, each associated with a batch of Z_b contiguous axial slices. This allows us to leverage local axial context while maintaining efficient parallelization. Each subvolume, defined over (x, y, z_k, t), is reconstructed independently by a dedicated INR.
Given a batch of temporal sinogram measurements

$$\mathbf{Y} = [\mathbf{Y}_0, \ldots, \mathbf{Y}_{T-1}], \quad \text{with} \quad \mathbf{Y}_t \in \mathbb{R}^{N_d \times N_\theta \times Z_b}, \tag{29}$$

where T is the number of temporal acquisitions, N_d the number of detector elements, N_θ the number of projection angles, and Z_b the number of jointly reconstructed axial slices. The final optimization problem for each axial subvolume becomes:

$$\min_{\theta} \mathcal{L}\left(P_{\boldsymbol{\Theta}}\{\mathbf{X}\}, \mathbf{Y}\right) + \lambda_s \, \mathrm{TV}_{\mathrm{space}}(\mathbf{X}) + \lambda_a \, \mathrm{TV}_{\mathrm{axial}}(\mathbf{X}) + \lambda_t \, \mathrm{TV}_{\mathrm{time}}(\mathbf{X}), \tag{30}$$

where $P_{\boldsymbol{\Theta}}$ the forward projector is applied to the reconstructed volume $\mathbf{X}$, parameterized by the INR weights $\boldsymbol{\theta}$, and the TV terms promote smoothness across the spatial, axial, and temporal dimensions, respectively.

In this work, each batch is treated independently. However, alternative strategies may improve consis-

tency across the full volume. For example, one may adopt a sliding window approach where neighboring axial subvolumes overlap, and impose coherence through additional coupling terms. A simple regularization between overlapping regions $\Omega = \mathcal{W}_m \cap \mathcal{W}_n$ could be written as

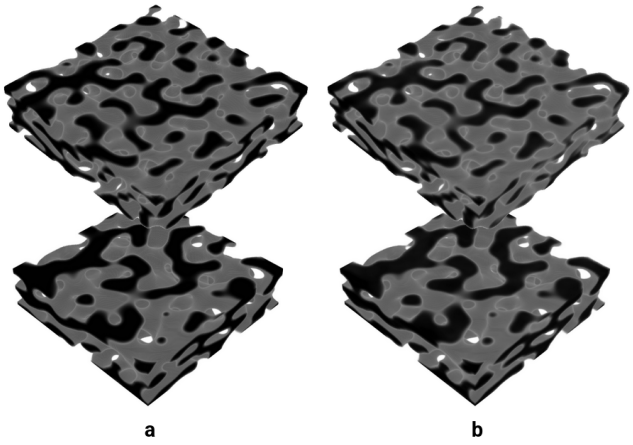
$$\lambda_c \sum_{z \in \Omega} \left\| x_z^{(m)} - x_z^{(n)} \right\|^2, \tag{31}$$

or, in a stricter consensus setting, by enforcing equality constraints of the form

$$x_z^{(m)} = x_z^{(n)}, \quad \forall z \in \Omega. \tag{32}$$

A practical limitation of our current architecture is that INCODE cannot directly handle 4D input ($N_d \times N_\theta \times Z_b \times T$). One option is to replace the pretrained 2D feature extractor with a 3D convolutional architecture (e.g., ResNet3D-18), at the cost of increased complexity. Alternatively, the input can be flattened so that each slice is processed independently using the original 2D network. We adopt this lightweight strategy for simplicity and memory efficiency.

Figure 8 illustrates reconstruction results for a 256×256×64 volume undergoing spinodal decomposition.



▼ **Figure 8**: 3DT reconstruction results of a $256 \times 256 \times 64$ volume undergoing spinodal decomposition. (a) Ground truth. (b) Our method. Both (a) and (b) use interlaced acquisition with $K = 16$ and $N_\theta = 256$, and a detector size of 363 to ensure full coverage. Top view: volume at the onset of the transformation; bottom view: volume at the end of the transformation.

7. Discussion

From an optimization perspective, our INR-based reconstruction method shares conceptual similarities with classical model-based iterative reconstruction (MBIR) approaches, as it incorporates a data fidelity term derived from the forward model, combined with explicit spatial and temporal regularization. In the case of TIMBIR, for example, the qGGMRF prior acts similarly to a total variation (TV) regularization [2]. However, unlike traditional MBIR techniques that operate on discretized image grids, our approach leverages a continuous and implicit neural representation, introducing a learned inductive bias and enabling flexible resolution handling as well as parameter sharing across the temporal dimension.

The model selection strategy used in our experiments, based on tracking the mean residual between primal and dual variables, $\mathbf{x} - \mathbf{q}$ is likely suboptimal. While this criterion promotes internal consistency within the ADMM framework, we observed that it can lead to slight degradations in PSNR and SSIM at later iterations, missing the best model. This suggests that the metrics reported here are conservative and that more principled model selection strategies should be explored in future work.

Finally, although our evaluation is conducted on simulated data, which provides a controlled ground truth for benchmarking, validation on real experimental acquisitions remains an important next step to assess robustness and practical applicability under experimental conditions.

8. Conclusion

This work demonstrates the feasibility and effectiveness of leveraging INRs for dynamic XCT reconstruction under interlaced acquisition schemes. The proposed model outperforms traditional baselines across a variety of scenarios, while offering a modular and flexible architecture capable of accommodating detector non-idealities such as ring artifacts. These results highlight the potential of INRs as a compact and versatile framework for tomographic imaging in complex, time-resolved settings, such as in situ monitoring of alloy solidification.

Future work will focus on improving computational efficiency and architectural parallelism to ensure scalability on large datasets, including support for multi-GPU and distributed execution environments.

9. Acknowledgement

This work is funded by MIT Startup and the Nuclear Engineering University Program with contract number"DE-NE009491. E.J. acknowledges the support from the John Hardwick Career Development Chair, and M.B. acknowledges support from the MIT-INSTN (CEA) internship program.

10. Data and code availability

The code and data used in this study are available at https://github.com/JossouResearchGroupMIT/Interlaced-INR-XCT.

References

[1] Amer Essakine, Yanqi Cheng, et al. Where do we stand with implicit neural representations? a technical and performance survey. *arXiv preprint arXiv:2411.03688*, 2024.

[2] K. Aditya Mohan, S. V. Venkatakrishnan, John W. Gibbs, Emine Begum Gulsoy, Xianghui Xiao, Marc De Graef, Peter W. Voorhees, and Charles A. Bouman. Timbir: A method for time-space reconstruction from interlaced views. *IEEE Transactions on Computational Imaging*, 1(2):96–111, 2015.

[3] Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. Space-time tomography for continuously deforming objects. *ACM Trans. Graph.*, 37(4), July 2018.

[4] Guangming Zang, Ramzi Idoughi, Ran Tao, Gilles Lubineau, Peter Wonka, and Wolfgang Heidrich. Warp-and-project tomography for rapidly deforming objects. *ACM Transactions on Graphics (TOG)*, 38(4), 2019.

[5] Yu Sun, Jiaming Liu, Mingyang Xie, Brendt Wohlberg, and Ulugbek S. Kamilov. Coil: Coordinate-based internal learning for imaging inverse problems, 2021.

[6] Guangming Zang, Ramzi Idoughi, Rui Li, Peter Wonka, and Wolfgang Heidrich. Intratomo: Self-supervised learning-based tomography via sinogram synthesis and prediction. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1940–1950, 2021.

[7] Kunal Gupta, Brendan Colvert, and Francisco Contijoch. Neural computed tomography, 2022.

[8] Ruyi Zha, Yanhao Zhang, and Hongdong Li. *NAF: Neural Attenuation Fields for Sparse-View CBCT Reconstruction*, page 442–452. Springer Nature Switzerland, 2022.

[9] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multiresolution hash encoding. In *ACM SIGGRAPH*, 2022.

[10] Albert W. Reed, Hyojin Kim, Rushil Anirudh, K. Aditya Mohan, Kyle Champley, Jingu Kang, and Suren Jayasuriya. Dynamic ct reconstruction from limited views with implicit neural representations and parametric motion fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2238–2248, 2021.

[11] Muge Du, Zhuozhao Zheng, Wenying Wang, Guotao Quan, Wuliang Shi, Le Shen, Li Zhang, Liang Li, Yinong Liu, and Yuxiang Xing. Nonperiodic dynamic ct reconstruction using backward-warping inr with regularization of diffeomorphism (bird), 2025.

[12] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[13] Boris Hanin and David Rolnick. Complexity of linear regions in deep networks. *arXiv preprint arXiv:1901.09021*, 2019.

[14] Matthew Tancik, Pratul P Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nalini Raghavan, Utkarsh Singhal, Jonathan T Barron, Ren Ng, and Angjoo Kanazawa. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, 2020.

[15] Amirhossein Kazerouni, Reza Azad, Alireza Hosseini, Dorit Merhof, and Ulas Bagci. Incode: Implicit neural conditioning with prior knowledge embeddings. *arXiv preprint arXiv:2310.18846*, 2023.

[16] Mahrokh Najaf and Gregory Ongie. Accelerated optimization of implicit neural representations for ct reconstruction. 2025.

[17] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, and Tony Yu. scikit-image: image processing in python. *PeerJ*, 2:e453, jun 2014.

[18] Matteo Ronchetti. Torchradon: Fast differentiable routines for computed tomography. *arXiv preprint arXiv:2009.14788*, 2020.

[19] Matteo Ravasi and Ivan Vasconcelos. Pylops – a linear-operator python library for large scale optimization. 2019.

[20] Avinash C. Kak and Malcolm Slaney. *Principles of Computerized Tomographic Imaging*. Society for Industrial and Applied Mathematics, 2001.

[21] G.R. Ramesh, N. Srinivasa, and K. Rajgopal. An algorithm for computing the discrete radon transform with some applications. In *Fourth IEEE Region 10 International Conference TENCON*, pages 78–81, 1989.

[22] Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis, 2020.

[23] John W Cahn. On spinodal decomposition. *Acta Metallurgica*, 9(9):795–801, 1961.

[24] John W. Cahn and John E. Hilliard. Free energy of a nonuniform system. i. interfacial free energy. *The Journal of Chemical Physics*, 28(2):258–267, 02 1958.

[25] Anna Guell Izard, Jens Bauer, Cameron Crook, Vladyslav Turlo, and Lorenzo Valdevit. Ultra-high energy absorption multifunctional spinodal nanoarchitectures. *Small*, 15(45):1903834, 2019.

[26] Carlos M. Portela, Abishek Vidyasagar, Sebastian Krödel, and Chiara Daraio. Extreme mechanical resilience of self-assembled nanolabyrinthine materials. *Proceedings of the National Academy of Sciences*, 117(11):5686–5693, 2020.

[27] Nian Yang, Ruiyuan Tian, and Fei Du. Polymer-assisted spinodal decomposition enabling a three-dimensional interconnected porous $na_{3.4}fe_{2.4}(po_4)_{1.4}p_2o_7$@c material for enhanced sodium-ion batteries. *Discover Electrochemistry*, 2(1):6, 2025.

[28] Changshin Jo, Bo Wen, Hyebin Jeong, Sul Ki Park, Yeonguk Son, and Michael De Volder. Spinodal decomposition method for structuring germanium–carbon li-ion battery anodes. *ACS Nano*, 17(9):8403–8410, 2023. PMID: 37067407.

[29] Tiankai Yao, Adrian R Wagner, Xiang Liu, Anter EI-Azab, Jason M Harp, Jian Gan, David H Hurley, Michael T Benson, and Lingfeng He. On spinodal-like phase decomposition in u–50zr alloy. *Materialia*, 9:100592, 2020.

[30] Tiankai Yao, Amrita Sen, Adrian Wagner, Fei Teng, Mukesh Bachhav, Anter EI-Azab, Daniel Murray, Jian Gan, David H. Hurley, Janelle P. Wharry, Michael T. Benson, and Lingfeng He. Understanding spinodal and binodal phase transformations in u-50zr. *Materialia*, 16:101092, 2021.

[31] Stephen DeWitt, Shiva Rudraraju, David Montiel, W. Beck Andrews, and Katsuyo Thornton. Prisms-pf: A general framework for phase-field modeling with a matrix-free finite element method. *npj Computational Materials*, 6(1), 03 2020.

[32] S. Fetni, J. Delahaye, L. Duchêne, A. Mertens, and A. Habraken. Adaptive time stepping approach for phase-field modeling of phase separation and precipitates coarsening in additive manufacturing alloys. In *COMPLAS 2021: XV International Conference on Computational Plasticity*, 2022. Accessed via Scipedia.

[33] J. H. Hubbell and S. M. Seltzer. Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients (version 1.4). http://physics.nist.gov/xaamdi, 2004. National Institute of Standards and Technology, Gaithersburg, MD. Accessed: 2025, September 25.

[34] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[35] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.

[36] Honghai Yu and Stefan Winkler. Image complexity and spatial information. In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, pages 12–17. IEEE, 2013.

[37] M. Rivers. Tutorial introduction to x-ray computed microtomography data processing. http://www.aps.anl.gov/xraytomography/tutorial.html, May 1998. Accessed September 19, 2025.

[38] Peter J. Huber. Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics*, 35(1):73 – 101, 1964.

[39] Paul W. Holland and Roy E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics - Theory and Methods*, 6(9):813–827, 1977.

[40] Bhaskar Mukhoty, Govind Gopakumar, Prateek Jain, and Purushottam Kar. Globally-convergent iteratively reweighted least squares for robust regression problems, 2020.

[41] Dianne P. O'Leary. Robust regression computation using iteratively reweighted least squares. *SIAM Journal on Matrix Analysis and Applications*, 11(3):466–480, 1990.

[42] Pierre Paleo and Alessandro Mirone. Ring artifacts correction in compressed sensing tomographic reconstruction. *Journal of Synchrotron Radiation*, 22:1268–1278, 07 2015.

[43] Kazumi Murata and Koichi Ogawa. Ring-artifact correction with total-variation regularization for material images in photon-counting ct. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(4):568–577, 2021.

[44] Qing Wu, Hongjiang Wei, Jing xin Yu, and Yuyao Zhang. Unsupervised multi-parameter inverse solving for reducing ring artifacts in 3d x-ray cbct. *ArXiv*, abs/2412.05853, 2024.