Physics-Guided Diffusion Priors for Multi-Slice Reconstruction in Scientific Imaging

Laurentius Valdy[1], Richard D. Paul[1,2], Alessio Quercia[1,3], Zhuo Cao[1], Xuan Zhao[1], Hanno Scharr[1], Arya Bangun[1]

[1]IAS-8, Forschungszentrum Jülich, Jülich, Germany
[2]Dept. of Statistics, LMU Munich, Munich, Germany
[3]Dept. of Computer Science, RWTH Aachen University, Aachen, Germany

Abstract

Accurate multi-slice reconstruction from limited measurement data is crucial to speed up the acquisition process in medical and scientific imaging. However, it remains challenging due to the ill-posed nature of the problem and the high computational and memory demands. We propose a framework that addresses these challenges by integrating partitioned diffusion priors with physics-based constraints. By doing so, we substantially reduce memory usage per GPU while preserving high reconstruction quality, outperforming both physics-only and full multi-slice reconstruction baselines for different modalities, namely Magnetic Resonance Imaging (MRI) and four-dimensional Scanning Transmission Electron Microscopy (4D-STEM). Additionally, we show that the proposed method improves in-distribution accuracy as well as strong generalization to out-of-distribution datasets.

Introduction

Reconstructing multi-slice images has become increasingly important in scientific disciplines, such as medical imaging, plant science, biology, and materials science (Midgley and Weyland 2003; Lee et al. 2023; Schulz et al. 2012; Chung et al. 2023; Bangun et al. 2025; Liu et al. 2024), mainly to enhance the visualization, analysis, and interpretation of the complex object structures. A persistent challenge in MRI is the inherently slow data acquisition process, which is further compounded by the growing demand for 3D object reconstruction through multi-slice (Zbontar et al. 2018; Bangun et al. 2025). Hence, the main challenge in accelerating MRI is to reconstruct multi-slice images from undersampled measurements. Similarly, in four-dimensional scanning transmission electron microscopy (4D-STEM) only intensity measurements are recorded in electron detector, thereby creating a demand to retrieve 3D phase information of the object under investigation, i.e., electrostatic potential of materials, through multi-slices model (Maiden, Humphry, and Rodenburg 2012; Bangun et al. 2022; Diederichs et al. 2024). Most proposed model-based methods for multi-slice reconstruction for both MRI and 4D-STEM, highly dependent on the initialization of the algorithm, yielding poor flexibility and often leading to suboptimal reconstructions.

A. Q. and R. D. P. were funded by the Helmholtz School for Data Science in Life, Earth, and Energy (HDS-LEE)

Diffusion models have become powerful tools for generating both 2D and 3D images, achieving notable success in computer vision, data augmentation, and image/video generation (Ho, Jain, and Abbeel 2020; Ho et al. 2022; Sohl-Dickstein et al. 2015; Song and Ermon 2019, 2020; Song et al. 2020). These models excel at capturing complex data distributions and generating high-fidelity images by iteratively refining samples from a random distribution into coherent structures. Despite their success with natural images, applying diffusion models to scientific images is challenging due to their unique characteristics, which arise from specialized modalities such as microscopy, spectroscopy, or medical imaging. The incorporation of multi-slice reconstruction requirements further amplifies these challenges. In particular, scientific data requires volumetric consistency, robustness to acquisition physics, and interpretability, which conventional diffusion frameworks do not provide directly. Therefore, applying diffusion models directly to scientific imaging raises questions about their suitability and whether they can enhance reconstruction quality while enabling fast image generation.

We propose a framework that integrates partitioned diffusion priors with physics-based forward models, providing scalable training and physics-consistent inference for multi-slice reconstruction, namely DART: Diffusion-Alternating Multi-slice Reconstruction Technique and DRIFT: Diffusion-Refined Initialization for Multi-slice Reconstruction. We empirically show that our algorithms are efficient in computation and memory usage (more than $8\times$) and robust to out-of-distribution (OOD) data.

Background

Diffusion Models

Diffusion models use a neural network $\mathcal{S}_\theta\left(\boldsymbol{X}_t, t\right)$ to learn original data $\boldsymbol{X}_0$, where training data is perturbed at each step t with variance β_t,

$$\boldsymbol{X}_t = \sqrt{1-\beta_t}\boldsymbol{X}_{t-1} + \sqrt{\beta}_t\boldsymbol{Z}_{t-1}, \;\; t \in [T],$$

with $\boldsymbol{Z}$ a Gaussian noise. The trained model with optimal parameters θ^* is used in inference via sampling (Ho, Jain,

Code: https://jugit.fz-juelich.de/ias-8/Distributed_3DDM

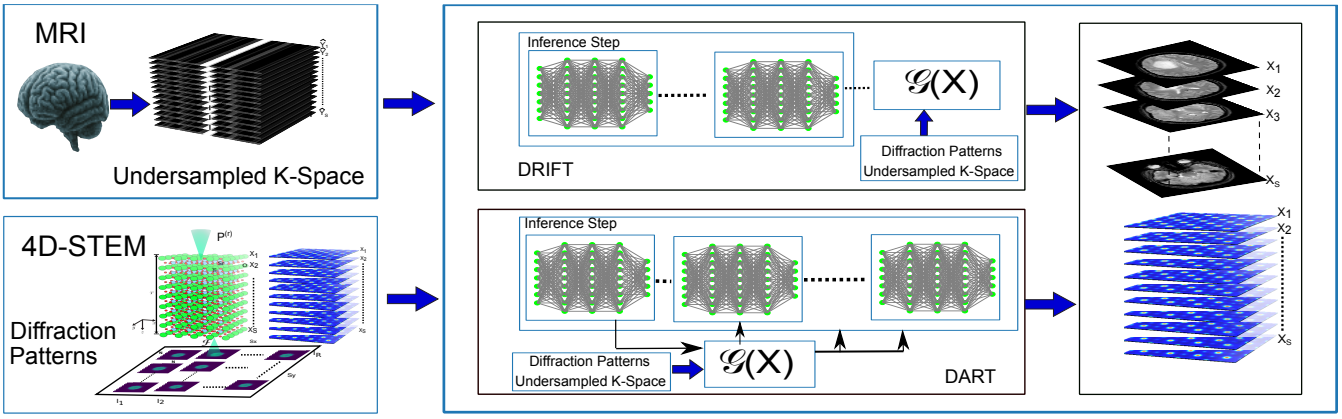

Figure 1: Data acquisition of a brain in MRI (k-space) and 4D-STEM (diffraction patterns) of crystalline materials, as well as proposed methods: DART (alternating update between trained diffusion prior and physics constraints) and DRIFT (diffusion priors as initialization before applying physical constraints). Physical constraints $\mathcal{G}(X)$ are adapted depending on the modality (MRI or 4D-STEM).

and Abbeel 2020; Ho et al. 2022)

$$\boldsymbol{X}_{t-1} = \frac{1}{1-\beta_t}\left(\boldsymbol{X}_t + \beta_t \mathcal{S}_\theta^*\left(\boldsymbol{X}_t, t\right)\right) + \sqrt{\beta_t}\mathbf{Z}_t. \quad (1)$$

Intuitively, diffusion models learn to iteratively denoise random noise into structured data samples. This makes diffusion models attractive as generative priors for scientific imaging.

Physics of MRI and 4D-STEM

In many physical modalities, due to accelerated data acquisition or physical limitations of the measurement process, only limited data can be acquired. For example, to accelerate 3D MRI acquisition, undersampled k-space data are collected and can be modeled as $\hat{\mathbf{Y}}_s = \mathbf{M} \circ \mathbf{Y}_s$, where $\mathbf{M}$ selects rows/columns from full k-space $\mathbf{Y}_s \in \mathbb{C}^{N\times N}$ for each slice $s \in \{1, 2, \dots S\}$.

A similar condition appears in multi-slice 4D-STEM, where only intensity measurement can be recorded. In multi-slice 4D-STEM, the exit wave after the first slice of crystalline materials $\mathbf{X}_1 \in \mathbb{C}^{N\times N}$ is given by

$$\mathbf{E}_1^{(r)} = \mathbf{X}_1 \circ \mathbf{P}^{(r)},$$

where $\mathbf{P}^{(r)}$ is the focused electron beam at scan point r in two-dimensional coordinate axes with element-wise product $\circ$. The exit wave is propagated through vacuum distance between slices using Fresnel operator $\mathcal{V}$,

$$\mathbf{E}_s^{(r)} = \mathbf{X}_s \circ \mathcal{V}\left(\mathbf{E}_{s-1}^{(r)}\right), s \in \{2, \dots, S\}.$$

At the last slice S, the exit wave $\mathbf{E}_S^{(r)}$ produces diffraction patterns with Fourier transform $\mathcal{F}$, recorded as

$$\mathbf{I}^{(r)} = \left|\mathcal{F}\left(\mathbf{E}_S^{(r)}\right)\right|^2 \in \mathbb{R}^{N\times N}.$$

The acquisition process for both undersampled MRI and 4D-STEM are shown in Figure 1.

The goal in both multi-slice MRI and 4D-STEM is to reconstruct multi-slice object $X \in \mathbb{C}^{S\times N\times N}$ given incomplete measurement data, i.e., undersampled k-space in MRI or phase retrieval problem from projection intensities with total scan $R = S_x \times S_y$ in 4D-STEM. The optimization problem for 4D-STEM in terms of Frobenius norm is

$$\boldsymbol{X}^* = \underset{\boldsymbol{X}\in\mathbb{C}^{S\times N\times N}}{\arg\min} \sum_{r=1}^{R} \left\| \mathbf{I}^{(r)} - \left|\mathcal{F}\left(\mathcal{H}\left(\boldsymbol{X}, \mathbf{P}^{(r)}\right)\right)\right|^2 \right\|_F^2 + \mathcal{R}\left(\boldsymbol{X}\right), \quad (2)$$

where $\mathcal{H}$ models slice interaction and Fresnel propagation in forward 4D-STEM multi-slice, $\mathbf{P}^{(r)}$ is the known focused probe, and $\mathcal{R}$ is a regularizer. For MRI, the corresponding optimization is

$$\boldsymbol{X}^* = \underset{\boldsymbol{X}\in\mathbb{C}^{S\times N\times N}}{\arg\min} \sum_{s=1}^{S} \left\| \hat{\mathbf{Y}}_s - \mathbf{M} \circ \left(\mathcal{F}\left(\mathbf{X}_s\right)\right) \right\|_F^2 + \mathcal{R}\left(\boldsymbol{X}\right), \quad (3)$$

In both settings, incomplete measurements (missing k-space lines in MRI, missing phase information in 4D-STEM) make recovery particularly challenging. Moreover, solutions to these optimization problems depend heavily on initialization, motivating the integration of diffusion priors trained on relevant data distributions. We denote optimization problems (2) (3) as $\mathcal{G}\left(\boldsymbol{X}\right)$.

Methods

We propose a physics-guided partitioned diffusion inference framework for multi-slice reconstruction by incorporating trained neural network to generate multi-slice priors. Since the training is straightforward by partitioning the multi-slice data of the MRI object, i.e., brain, and the electron microscopy object, i.e., crystalline materials, we focus on the inference aspect first and defer the training procedure to the numerical section.

Measurement data (MRI k-space or 4D-STEM diffraction patterns) are incorporated through the respective physical models, ensuring that inference respects instrument constraints. To scale inference across multiple GPUs, S slices

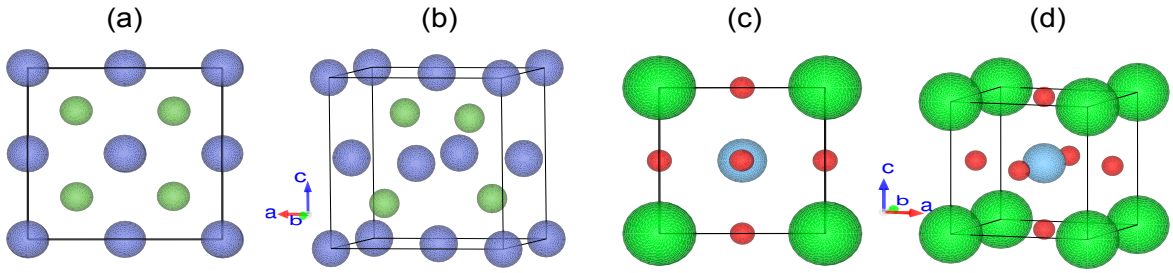

Figure 2: Example of cubic crystal materials: unit cells of gallium arsenide (GaAs) with volume dimension 5.6533^3Å^3 (a),(b) projection and 3D visualization, strontium titanate ($SrTiO_3$) with volume dimension 3.905^3Å^3 (c), (d) projection and 3D visualization. In the multi-slice method, each slice is obtained by calculating the projection of each atomic plane along the z direction.

Algorithm 1: Physics-Guided Partitioned Diffusion Inference (DRIFT / DART)

Require: #GPUs G, model $\mathcal{G}$, measurement data $\hat{\mathbf{Y}}_s$ (MRI) or $\mathbf{I}^{(r)}$ (4D-STEM)

Initialize $\mathbf{X} = \begin{cases} \mathbf{X} \in \mathbb{C}^{L \times S \times N \times N} & \text{(DRIFT)} \\ \mathbf{X} \in \mathbb{C}^{S \times N \times N} & \text{(DART)} \end{cases}$

for $t = T$ **to** 1 **do**
  Slice size $B_g = \lceil S/G \rceil$
  for all GPU $g = 0 \dots G-1$ in parallel **do**
    Block $S_g = \min((g+1)B_g, S) - gB_g$
    Extract $\mathbf{X}_g \in \mathbb{C}^{S_g \times N \times N}$ or
      $\mathbf{X}_g \in \mathbb{C}^{L \times S_g \times N \times N}$
    $\mathbf{X}_g \leftarrow \texttt{DDPMSampler}(\mathbf{X}_g, t)$ in (1)
    Send $\mathbf{X}_g \rightarrow$ root
  end for
  Gather all $\mathbf{X}_g \rightarrow \mathbf{X}$
  if mode = DART **then**
    $\mathbf{X} \leftarrow \mathcal{G}(\hat{\mathbf{Y}}_s \text{ or } \mathbf{I}^{(r)}, \mathbf{X})$ ▷ apply physics model
  end if
end for
if mode = DRIFT **then**
  $\mathbf{X} \leftarrow \underset{l \in [L]}{\text{argmax}}\ \texttt{SSIM}(\hat{\mathbf{Y}}_s \text{ or } \mathbf{I}^{(r)}, \mathbf{X})$
  $\mathbf{X} \leftarrow \mathcal{G}(\hat{\mathbf{Y}}_s \text{ or } \mathbf{I}^{(r)}, \mathbf{X})$ ▷ physics refinement
end if
return $\mathbf{X} \in \mathbb{C}^{S \times N \times N}$

are distributed uniformly over G GPUs ($\lceil S/G \rceil$ each), enabling parallel processing and efficient memory usage. Our approach includes two complementary variants:

- **DART** (Diffusion-Alternating Multi-slice Reconstruction): alternates between diffusion prior updates and physics-based constraints at each iteration.
- **DRIFT** (Diffusion-Refined Initialization for Multi-slice Reconstruction): uses diffusion priors to generate a high-quality initialization, followed by physics-based refinement.

The inference procedure is summarized in Algorithm 1, as well as in Figure 1 where the main difference between DART and DRIFT is visualized. In DRIFT model, the SSIM function is calculated to measure the similarity of the initialization with the measurement data. For MRI, the SSIM is evaluated after inverse Fourier transform of undersampled k-space $\hat{\mathbf{Y}}_s$ per slice, i.e., $\mathcal{F}^{-1}\left(\hat{\mathbf{Y}}_s\right)$. On the contrary, in 4D-STEM, the bright field image is calculated by integrating or summing the detector image to have two-dimensional data related to the scan dimension before calculating SSIM, i.e., sum over last two axes of dimension (S_y, S_x, N, N) to get (S_y, S_x).

The operator $\mathcal{G}(\mathbf{X})$ enforces measurement consistency with the underlying physical model. For MRI, this corresponds to a *data-consistency step* that replaces the sampled entries in k-space with the acquired measurements:

$$\mathbf{X}_s^{(k+1)} = \text{Prox}\left(\mathbf{X}_s^{(k)} - \lambda \nabla_{\mathbf{X_s}} \sum_{s=1}^{S} \left\| \hat{\mathbf{Y}}_s - \mathbf{M} \circ (\mathcal{F}(\mathbf{X}_s)) \right\|_F^2\right), \tag{4}$$

where $\mathcal{F}$ denotes the Fourier transform, $\mathbf{M}$ is the sampling mask, and λ is a step size.

For 4D-STEM, $\mathcal{G}$ corresponds to one step of gradient descent on the intensity-based loss in Eq. (2):

$$X^{(k+1)} = \text{Prox}\left(X^{(k)} - \eta \nabla_X \sum_{r=1}^{R} \left\| \mathbf{I}^{(r)} - \left| \mathcal{F}(\mathcal{H}(X^{(k)}, \mathbf{P}^{(r)})) \right|^2 \right\|_F^2\right), \tag{5}$$

where η is the step size and the strategy to choose the step size is given for instance in (Xu et al. 2018). The function Prox is the proximal function for regularization, for instance, soft thresholding for the sparse structure of the data (Parikh, Boyd et al. 2014).

In practice, we use a fixed number of physics-update iterations per diffusion step. For instance, in DART approach for every diffusion step we perform one step gradient update. In DRIFT, we perform 100 steps gradient update after initialization from the diffusion model for complete step T. This modular design allows the same framework to generalize across different modalities by substituting the appropriate physical constraint $\mathcal{G}$.

Numerical Experiments

We evaluate our methods on two scientific imaging applications: MRI and 4D-STEM. Apart from the contribution for

inference algorithms, DART and DRIFT, diffusion model is also trained with multi-slice MRI and crystalline materials data.

Training Procedures

For MRI, we use T2-FLAIR Brain Tumor Segmentation (BraTS) 2020 (Bakas et al. 2017, 2018; Menze et al. 2014) with dimensions $155 \times 240 \times 240$. For 4D-STEM, we use a cubic crystal system with lattice structure from the Materials Project (Jain et al. 2013). Multi-slice algorithm implemented in (Durham et al. 2022) is used to generate slices of atomic potentials. The dataset contains pixel dimensions $10 \times 80 \times 80$ per file with a total of 3533 files. A visualization of the material can be seen in Figure 2. We use a video diffusion model (Ho et al. 2022) that extends 2D U-Nets using convolutions and attention mechanisms that operate separately over spatial and temporal dimensions, to efficiently capture both the appearance and motion present in video data, which is aligned with both multi-slice MRI and crystalline materials. The model is used with base width 64 and channel multipliers (1, 2, 4, 8). Additionally, it is trained with $700{,}000$ steps using ℓ_1 loss, learning rate 10^{-4}, batch-level gradient accumulation over 2 steps, and other hyperparameters following (Ho et al. 2022). To address the computational complexity of training, we adopt a slice-wise training strategy. The axial slices of MRI and crystal data are partitioned into G groups with respect to the number of GPUs and each assigned to an independent diffusion model. To guarantee the reconstruction of each measurement data for MRI (k-space) and 4D-STEM (diffraction patterns), it should be incorporated in the inference process, as presented in Figure 1.

Magnetic Resonance Imaging

We report numerical results in terms of SSIM in Table 1 evaluated on 30 random multi-slice data. Figure 3 shows the slice reconstruction for various algorithms. It can be seen that DART achieves higher SSIM values for reconstructing slices and preserves anatomical details close to the ground truth despite the undersampled k-space using a uniform mask. Overall, these results demonstrate that both proposed methods (DART and DRIFT) successfully reconstruct the structural information, closely matching the fully sampled ground truth data despite undersampled k-space.

4D-Scanning Transmission Electron Microscopy

For evaluation, 30 multi-slices ground truth crystalline materials are generated. The intensity of diffraction patterns is acquired with total scan $R = R_x \times R_y = 80 \times 80$ with detector's dimension 80×80. Table 1 presents a comparison of SSIM values across other methods, namely Sparse Decomposition (Bangun et al. 2022), 3PIE (Maiden, Humphry, and Rodenburg 2012), and Torchslice (Diederichs et al. 2024). The phase projection of materials is used to evaluate the SSIM. DRIFT achieves the highest mean SSIM, suggesting that it delivers the most accurate reconstruction quality by leveraging a trained diffusion model before generating promising initial guesses and refining them with a physics-based iterative solver. One of the reasons DART performs worse in 4D-STEM case is that, unlike the MRI case, where for each slice we have a unique pair of k-space images, in 4D-STEM, the measurement data is only the projected intensity, which leads to slow convergence. Figure 3 presents 2D phase projections for two cubic crystalline materials, namely $CoPt_3$ and Tb_3InC. Each projection is annotated with its SSIM value, reflecting the structural similarity to the ground truth. SOTA algorithms also achieve high SSIM for this dataset but lower SSIM on average.

Out-of-Distribution Data

We additionally validate our methods against baselines on out-of-distribution (OOD) data, which is critical for real-world deployment when large annotated datasets are not available. In Table 2, we present the evaluation results on the plant roots dataset (Schulz et al. 2012) and hexagonal crystalline materials from Material Projects (Jain et al. 2013). Both DART and DRIFT perform reconstruction using models pre-trained on MRI BraTS and cubic crystal data, with an $8\times$ acceleration factor mask based on a Gaussian distribution to undersample MRI measurement data (k-space). DART and DRIFT maintain higher SSIM under OOD settings for both MRI and 4D-STEM, respectively, indicating better generalization. Figure 4 shows example reconstructions, where both DART and DRIFT further preserve both structural and intensity information, even when tested on out-of-distribution data.

Ablation Studies

Incorporating diffusion models into scientific imaging requires consideration of the physical characteristics of the data acquisition process.

Figure 5 shows the trade-off between inference time and the quality of reconstruction given measurement data for vanilla diffusion model and the proposed algorithms across 8 GPUs. In the absence of physical constraints and measurement data, the vanilla diffusion model fails to produce reconstructions that are faithful to the ground truth, as indicated by lower SSIM. It can be observed that DART increases a small fraction of the run time for inference step $T \in \{10, 100, 1000\}$. However, for DRIFT, the run time highly depends on the number of images we generate.

We also show that distributing the data to more GPUs helps to speed up the inference process and reduce memory allocation. The memory demand for the vanilla diffusion model with multi-slice dimension $155 \times 240 \times 240$ is very high, i.e., 54.55 GB. This might exceed many common GPU configurations and lead to out-of-memory issues. Therefore, partitioning the job across multiple GPUs not only produces a speed up but also helps to keep per-GPU memory usage within our limits.

Quantitative Evaluation of Generated Multi-Slice

Beyond time and memory benchmarks, we assess the distributional similarity between multi-slice images generated from a vanilla partitioned diffusion model and ground-truth using FVD (Fréchet Video Distance) (Unterthiner et al. 2019; Ge et al. 2024) and JEDi (JEPA Embedding Distance)

Table 1: Mean and standard deviation of SSIM from BraTS data and cubic crystal data. The **bold** and underline represent the best and second-best results.

Dataset	Mask Types	Methods	SSIM (↑)
BRATS MRI	Uniform $2\times,0.15$	DART	**0.968** $\pm$ 0.011
		DRIFT ($L = 16$)	<u>0.938</u> $\pm$ 0.023
		Projection-Based (Bangun et al. 2025)	0.844 $\pm$ 0.027
		CS MRI (Lustig, Donoho, and Pauly 2007)	0.804 $\pm$ 0.025
		TV (Block, Uecker, and Frahm 2007)	0.803 $\pm$ 0.024
Cubic Crystal Data	–	DART	0.607 $\pm$ 0.406
		DRIFT ($L = 16$)	**0.899** $\pm$ 0.091
		Sparse Decom (Bangun et al. 2022)	0.831 $\pm$ 0.198
		3PIE (Maiden, Humphry, and Rodenburg 2012)	<u>0.887</u> $\pm$ 0.243
		Torchslice (Diederichs et al. 2024)	0.823 $\pm$ 0.115

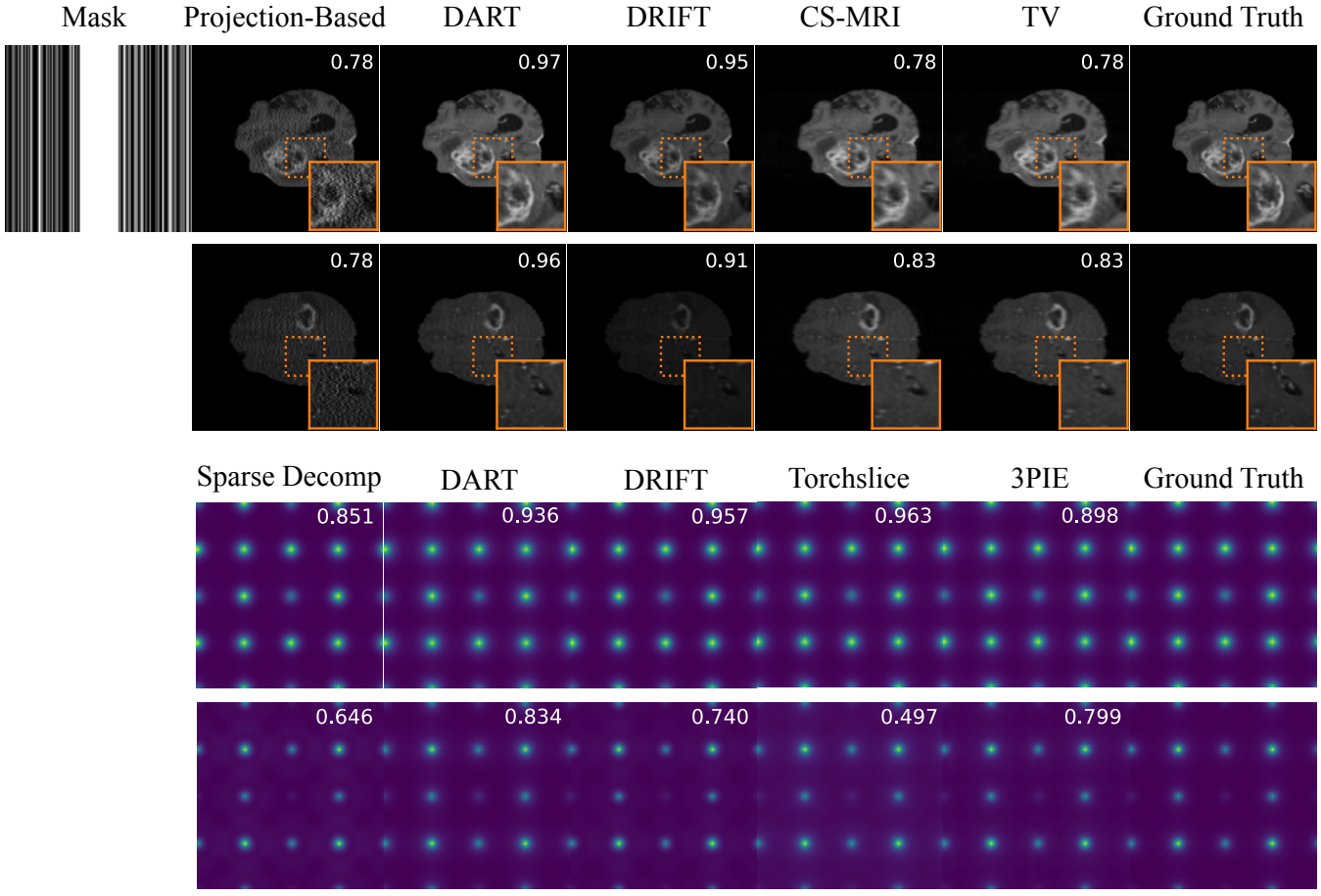

Figure 3: Single slice MRI with zoomed-in region of interest from the volume reconstruction of file BraTS20 Training 338 t1ce (top) and BraTS20 Training 039 t1ce (bottom); Projection-based (Bangun et al. 2025); DART; DRIFT; CS MRI (Lustig, Donoho, and Pauly 2007); Total Variation (Block, Uecker, and Frahm 2007). Phase Projection of crystalline materials $CoPt_3$ (top) and Tb_3InC (bottom) benchmarking with Sparse Decomposition (Bangun et al. 2022); DART; DRIFT; Torchslice (Diederichs et al. 2024); 3PIE (Maiden, Humphry, and Rodenburg 2012).Top right are visual quality metrics, namely SSIM.

(Luo et al. 2025). In the absence of physics-based conditioning, a vanilla diffusion model stochastically generates multi-slice from the learned prior, so distribution-level metrics such as FVD and JEDi are the appropriate measure. We report three variants: FVD with i3D features, FVD with VideoMAE features, and JEDi with V-JEPA features. Evaluations span 50–366 generated multi-slice data under two configurations: (i) single-GPU generation of $40 \times 240 \times 240$ volumes; (ii) 2-GPU distributed generation that produces two independent $20 \times 240 \times 240$ stacks and concatenates them

Table 2: Mean and standard deviation of SSIM from OOD data, namely Roots and Hexagonal Crystal Data. The **bold** and underline represent the best and second-best results.

Dataset	Mask Types	Methods	SSIM (↑)
Roots MRI	Gaussian $8\times,0.08$	DART	**0.813** $\pm$ 0.130
		DRIFT ($L = 16$)	<u>0.750</u> $\pm$ 0.174
		Projection-Based (Bangun et al. 2025)	0.611 $\pm$ 0.169
		CS MRI (Lustig, Donoho, and Pauly 2007)	0.676 $\pm$ 0.151
		TV (Block, Uecker, and Frahm 2007)	0.577 $\pm$ 0.163
Hexa Crystal Data	–	DART	0.588 $\pm$ 0.355
		DRIFT ($L = 16$)	**0.981** $\pm$ 0.010
		Sparse Decom (Bangun et al. 2022)	<u>0.977</u> $\pm$ 0.011
		3PIE (Maiden, Humphry, and Rodenburg 2012)	0.953 $\pm$ 0.041
		Torchslice (Diederichs et al. 2024)	0.573 $\pm$ 0.025

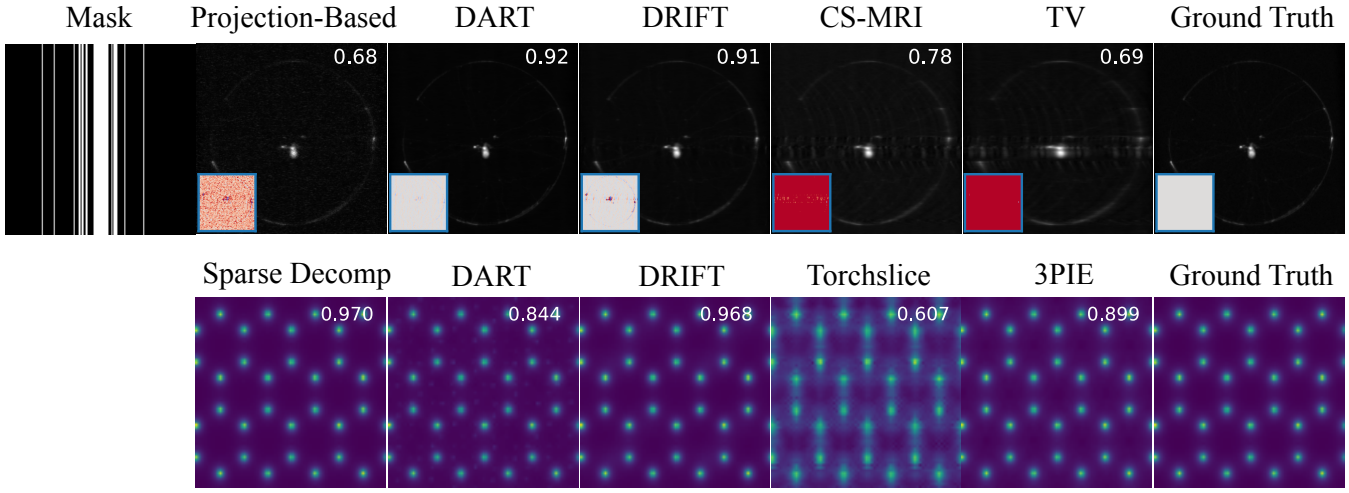

Figure 4: Single slice MRI from the volume reconstruction of file soybean roots; Projection-based (Bangun et al. 2025); DART; DRIFT; CS MRI (Lustig, Donoho, and Pauly 2007); Total Variation (Block, Uecker, and Frahm 2007). Phase Projection of crystalline materials WSe_2 benchmarking with Sparse Decomposition (Bangun et al. 2022); DART; DRIFT; Torchslice (Diederichs et al. 2024); 3PIE (Maiden, Humphry, and Rodenburg 2012). Top right are visual quality metrics, namely SSIM.

# Generated	**FVD-i3D** (↓)		**FVD-VideoMAE** (↓)		**JEDi** (↓)	
	3D (single)	**2-GPU (stack)**	**3D (single)**	**2-GPU (stack)**	**3D (single)**	**2-GPU (stack)**
50	2,099.01	2,638.97	251.70	751.61	7.74	9.64
100	2,109.51	2,675.64	261.79	762.91	7.49	9.54
150	2,187.89	2,653.14	263.68	759.70	7.36	10.06
200	2,051.56	2,653.39	257.39	753.80	7.76	9.92
250	2,036.96	2,626.47	256.36	749.66	7.78	9.81
300	2,018.10	2,584.11	252.91	750.68	7.71	9.67
366	2,034.72	2,615.10	254.56	754.38	7.76	9.79

Table 3: FVD and JEDi for MRI multi-slice generated data from vanilla diffusion model. 3D (single) denotes single-GPU generation of $40 \times 240 \times 240$. 2-GPU (stack) denotes distributed generation of two $20 \times 240 \times 240$ stacks concatenated to $40 \times 240 \times 240$. Lower is better.

into $40 \times 240 \times 240$.

Table 3 shows that the single-GPU configuration consistently outperforms the 2-GPU stacked setup across all sample sizes. The gap is largest for FVD with VideoMAE features, moderate for FVD-i3D, and smallest for JEDi, indicating that concatenating independently generated 20-slice stacks degrades space–time coherence captured by FVD, while JEDi is comparatively robust. Although the distributed

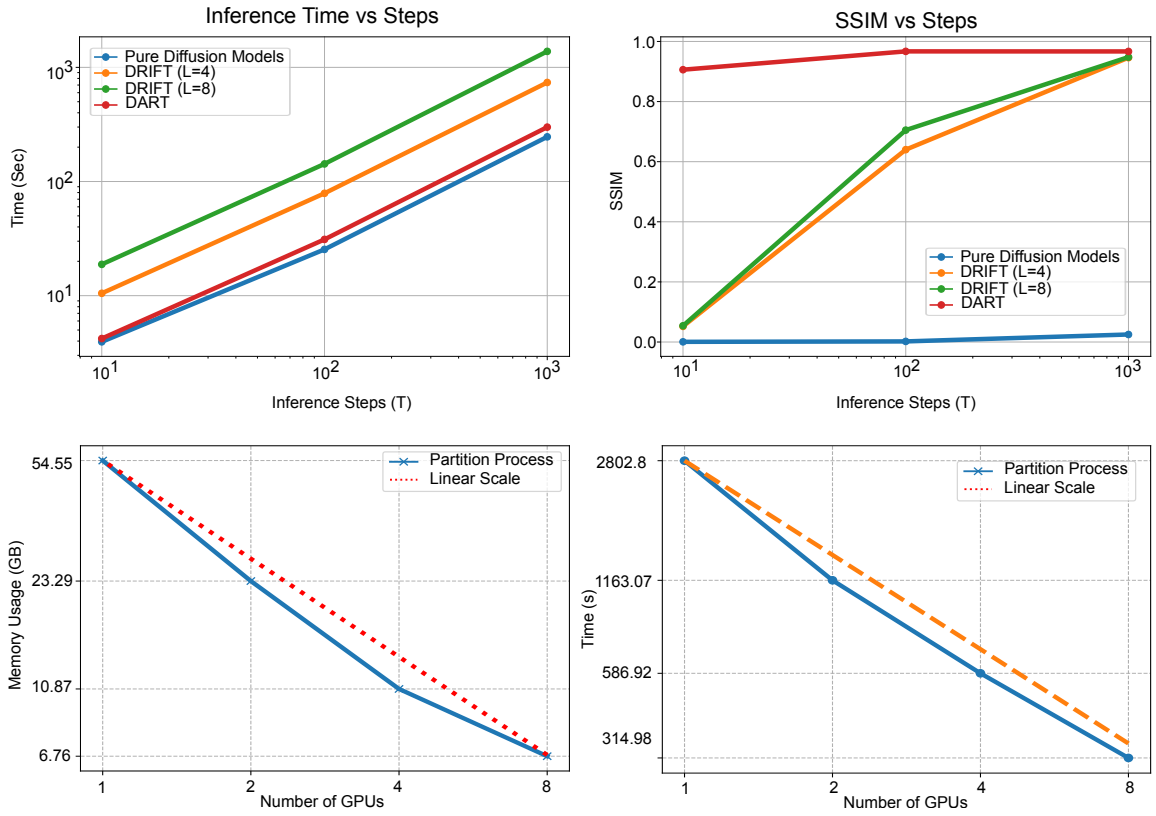

Figure 5: Top row: run time and SSIM of pure (vanilla) diffusion models (Ho et al. 2022), DART, and DRIFT. The experiments are conducted on a distributed process with 8 GPUs for multi-slice MRI data with dimension $155 \times 240 \times 240$. Bottom row: time and memory benchmarking for DART using various numbers of GPUs.

strategy performs worse, the absolute differences are limited, suggesting it remains viable under resource constraints. However, as we show in Figure 5, a vanilla 3D diffusion model without physics-based constraints (e.g., measurement models such as k-space in MRI or diffraction patterns in 4D-STEM) can yield non-unique reconstructions for exact ground truth data.

Conclusion

This work introduces a physics-guided framework that integrates partitioned diffusion priors for multi-slice reconstruction in scientific imaging. In particular, we propose two algorithms—DART: Diffusion–Alternating Multi-slice Reconstruction Technique and DRIFT: Diffusion–Refined Initialization for Multi-slice Reconstruction. By coupling learned diffusion priors with explicit forward models (e.g., k-space for MRI and diffraction patterns for 4D-STEM), our framework provides a principled bridge between data-driven generative modeling and physics-based reconstruction, yielding outputs that are both perceptually plausible and physically consistent. The proposed methods are not only computationally and memory efficient, but also effective in both MRI and 4D-STEM applications, especially for general real-time multi-slice reconstruction in scientific imaging applications. Potential directions include training and evaluation on experimental 4D-STEM diffraction patterns.

Acknowledgment

The authors gratefully acknowledge computing time on the supercomputer JURECA(Thörnig 2021) at Forschungszentrum Jülich under grant `delia-mp`.

References

Bakas, S.; Akbari, H.; Sotiras, A.; Bilello, M.; Rozycki, M.; Kirby, J. S.; Freymann, J. B.; Farahani, K.; and Davatzikos, C. 2017. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1): 1–13.

Bakas, S.; Reyes, M.; Jakab, A.; Bauer, S.; Rempfler, M.; Crimi, A.; Shinohara, R.; Berger, C.; Ha, S.; Rozycki, M.; et al. 2018. Identifying the best machine learning algorithms for brain tumor segmentation. *progression assessment, and overall survival prediction in the BRATS challenge*, 10.

Bangun, A.; Cao, Z.; Quercia, A.; Scharr, H.; and Pfaehler, E. 2025. MRI Reconstruction with Regularized 3D Diffusion Model (R3DM). In *Proceedings of the IEEE/CVF WACV*.

Bangun, A.; Melnyk, O.; März, B.; Diederichs, B.; Clausen, A.; Weber, D.; Filbir, F.; and Müller-Caspary, K. 2022. Inverse multislice ptychography by layer-wise optimisation

and sparse matrix decomposition. *IEEE Trans. on Comp. Imaging*, 8: 996–1011.

Block, K. T.; Uecker, M.; and Frahm, J. 2007. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 57(6): 1086–1098.

Chung, H.; Ryu, D.; McCann, M. T.; Klasky, M. L.; and Ye, J. C. 2023. Solving 3d inverse problems using pre-trained 2d diffusion models. In *Proceedings of the IEEE/CVF CVPR*, 22542–22551.

Diederichs, B.; Herdegen, Z.; Strauch, A.; Filbir, F.; and Müller-Caspary, K. 2024. Exact inversion of partially coherent dynamical electron scattering for picometric structure retrieval. *Nature Comm.*, 15(1): 101.

Durham, D. B.; Ophus, C.; Siddiqui, K. M.; Minor, A. M.; and Filippetto, D. 2022. Accurate quantification of lattice temperature dynamics from ultrafast electron diffraction of single-crystal films using dynamical scattering simulations. *Structural Dynamics*, 9(6).

Ge, S.; Mahapatra, A.; Parmar, G.; Zhu, J.-Y.; and Huang, J.-B. 2024. On the Content Bias in Frechet Video Distance. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 7277–7288.

Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *NeurIPS*, 33: 6840–6851.

Ho, J.; Salimans, T.; Gritsenko, A.; Chan, W.; Norouzi, M.; and Fleet, D. J. 2022. Video Diffusion Models. *ICLR 2022 Workshop Deep Generative Models for Highly Structured Datas*.

Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; et al. 2013. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1).

Lee, J.; Lee, M.; Park, Y.; Ophus, C.; and Yang, Y. 2023. Multislice electron tomography using four-dimensional scanning transmission electron microscopy. *Physical Review Applied*, 19(5): 054062.

Liu, X.; Rüttgers, M.; Quercia, A.; Egele, R.; Pfaehler, E.; Shende, R.; Aach, M.; Schröder, W.; Balaprakash, P.; and Lintermann, A. 2024. Refining computer tomography data with super-resolution networks to increase the accuracy of respiratory flow simulations. *Future Generation Computer Systems*, 159: 474–488.

Luo, G. Y.; Favero, G. M.; Luo, Z. H.; Jolicoeur-Martineau, A.; and Pal, C. 2025. Beyond FVD: Enhanced Evaluation Metrics for Video Generation Quality. *International Conference on Learning Representations (ICLR)*.

Lustig, M.; Donoho, D.; and Pauly, J. M. 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 58(6): 1182–1195.

Maiden, A. M.; Humphry, M. J.; and Rodenburg, J. M. 2012. Ptychographic transmission microscopy in three dimensions using a multi-slice approach. *JOSA A*, 29(8): 1606–1614.

Menze, B. H.; Jakab, A.; Bauer, S.; Kalpathy-Cramer, J.; Farahani, K.; Kirby, J.; Burren, Y.; Porz, N.; Slotboom, J.; Wiest, R.; et al. 2014. The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE trans. on medical imaging*, 34(10): 1993–2024.

Midgley, P. A.; and Weyland, M. 2003. 3D electron microscopy in the physical sciences: the development of Z-contrast and EFTEM tomography. *Ultramicroscopy*, 96(3-4): 413–431.

Parikh, N.; Boyd, S.; et al. 2014. Proximal algorithms. *Foundations and trends® in Optimization*, 1(3): 127–239.

Schulz, H.; Postma, J. A.; van Dusschoten, D.; Scharr, H.; and Behnke, S. 2012. 3D reconstruction of plant roots from MRI images. In *International Conference on Computer Vision Theory and Applications*, volume 2, 24–33. SciTePress.

Sohl-Dickstein, J.; Weiss, E.; Maheswaranathan, N.; and Ganguli, S. 2015. Deep unsupervised learning using nonequilibrium thermodynamics. In *ICML*, 2256–2265. PMLR.

Song, Y.; and Ermon, S. 2019. Generative modeling by estimating gradients of the data distribution. *NeurIPS 32*.

Song, Y.; and Ermon, S. 2020. Improved Techniques for Training Score-Based Generative Models. In *NeurIPS 33*.

Song, Y.; Sohl-Dickstein, J.; Kingma, D. P.; Kumar, A.; Ermon, S.; and Poole, B. 2020. Score-Based Generative Modeling through Stochastic Differential Equations. In *ICLR*.

Thörnig, P. 2021. JURECA: Data centric and booster modules implementing the modular supercomputing architecture at jülich supercomputing centre. *Journal of large-scale research facilities JLSRF*, 7: A182–A182.

Unterthiner, T.; van Steenkiste, S.; Kurach, K.; Marinier, R.; Michalski, M.; and Gelly, S. 2019. FVD: A new metric for video generation. *International Conference on Learning Representations (ICLR) Workshop DeepGenStruct*.

Xu, R.; Soltanolkotabi, M.; Haldar, J. P.; Unglaub, W.; Zusman, J.; Levi, A. F.; and Leahy, R. M. 2018. Accelerated Wirtinger flow: A fast algorithm for ptychography. *arXiv preprint arXiv:1806.05546*.

Zbontar, J.; Knoll, F.; Sriram, A.; Murrell, T.; Huang, Z.; Muckley, M. J.; Defazio, A.; Stern, R.; Johnson, P.; Bruno, M.; et al. 2018. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv:1811.08839*.