Space-Time Implicit Neural Representations for Atomic Electron Tomography on Dynamic Samples

Tiffany Chien
University of California, Berkeley
Berkeley, CA 94720
tiffany_chien@berkeley.edu

Colin Ophus
Lawrence Berkeley National Laboratory
Berkeley, CA 94720

Laura Waller
University of California, Berkeley
Berkeley, CA 94720

Abstract

Solving for the 3D atomic structure of unknown materials is a key problem in materials science. Atomic electron tomography (AET) is a technique capable of reconstructing the 3D position and chemical species of all atoms in a nanoscale sample from a series of 2D projections from different angles. One challenge in AET is carbon contamination that accumulates on the sample while collecting the tomographic projections, creating an unwanted temporal dynamic that degrades reconstruction quality when existing tomography algorithms expect a static sample. In this work, we use an unsupervised implicit neural representation (INR) as a space-time model to computationally remove the contamination and recover a clean 3D reconstruction, and show promising preliminary results on simulated data.

1 Introduction

Key questions in the field of materials science revolve around determining the properties and functionality of engineered materials, and solving the atomic structure of a material is one crucial step to understanding it at a fundamental level. Electron microscopy is the main tool for studying atomic structure because it achieves resolution on the scale of individual atoms. In this paper, we work with ptychographic atomic electron tomography (PAET), a method that can resolve 3D atomic structures with both light and heavy atoms. PAET uses ptychography to reconstruct the complex-field (phase and amplitude) of the sample from intensity measurements [1]. By rotating the sample and repeating the ptychographic procedure, we obtain a set of 2D phase projections of the sample from different angles, as shown in Fig. 1a. PAET then uses these projections to reconstruct the 3D atomic potential, which can be used to identify atoms and solve for the atomic structure of the material [2].

Unfortunately, during the time it takes to capture projections for PAET, stray carbon atoms will be attracted to the sample because the electron beam statically charges its surface. This accumulation of carbon contamination causes the sample to change over time, with different projection angles seeing different amounts of carbon, as shown in Figure 1b where a substantial amount of amorphous carbon can be seen growing on the surface of a carbon nanotube sample in experiment. Because existing tomography algorithms assume that the sample is static, this causes major reconstruction artifacts that make it impossible to fully solve the sample's atomic structure. Contamination is a factor that affects nearly every PAET experiment, and often results in unused datasets.

In this paper, we propose a space-time algorithm to solve for the dynamic carbon contamination jointly with the 3D reconstruction of the sample, such that the contamination can be computationally removed.

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

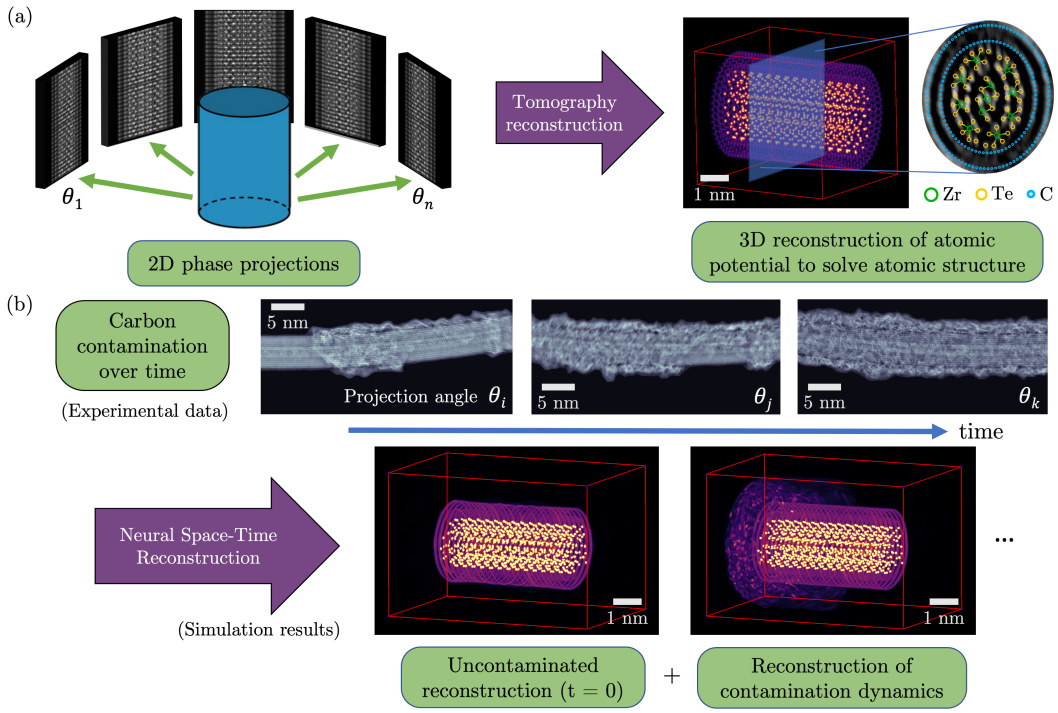

Figure 1: (a) Ptychographic atomic electron tomography (PAET) uses 2D phase projections at many different rotation angles (θ) to tomographically reconstruct the 3D atomic potential of a sample with single-atom resolution. (b) During the data acquisition process, carbon contamination accumulates on the sample, shown here in experimental data. This causes each projection to be a measurement of a slightly different sample, which degrades reconstruction quality if we assume the sample is static. Our neural space-time method reconstructs a full 3D video of the dynamic sample, such that we can computationally remove the contamination (results shown on simulated data).

Even though this joint optimization is an underdetermined problem, space-time algorithms take advantage of the redundant information shared between measurements taken at different times through spatiotemporal priors and regularization. Examples of space-time algorithms include modeling sample motion with smooth deformation fields between time points [3] and restricting the complexity of local motion with low-rank constraints [4]. However, our carbon contamination problem does not fit well into these existing models because new material appears on the sample over time. Here, we replace matrix-based representations of the reconstruction with an implicit neural representation (INR), a neural network trained to output the 3D atomic potential at a given input coordinate (x, y, z, t). Instead of using explicit regularizers, the network inherently mixes information from all time points and implicitly regularizes the reconstruction in space and time. This approach is inspired by work on view synthesis for dynamic scenes [5, 6]. It was recently applied to structured illumination super-resolution microscopy [7], and here we apply it to atomic tomography.

2 Methods

Our implicit neural representation (INR) serves as a continuous, implicit representation of the reconstructed 3D video by taking in space-time coordinates (x, y, z, t) and outputting the scalar value of the 3D atomic potential at each input coordinate. The input coordinates are encoded with a sinusoidal input encoding to improve the network's representation of high frequencies [8]. To get back an explicit matrix representation of the reconstruction, we need only query the network with all the 3D + time coordinates we are interested in.

As shown in Fig. 2, the optimization of the INR follows traditional gradient-based reconstruction methods, with the only difference being that the estimated reconstruction is represented by a network instead of a matrix. At each iteration, the network's estimate of the 3D reconstruction at each time

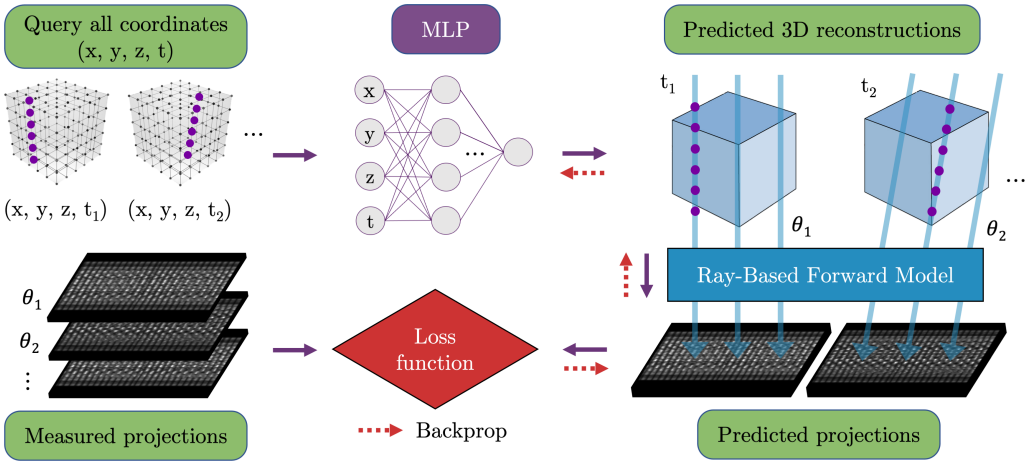

Figure 2: A single coordinate-based multi-layer perceptron (MLP) represents the 3D video reconstruction by taking in (x, y, z, t) coordinates and outputting the value of the 3D atomic potential at that coordinate. The network is optimized by querying it for its estimated 3D video reconstruction, passing the estimates through a ray-based tomography forward model, and comparing with measured projections. The gradient of the loss function is then backpropagated (red dashed lines) to update the network weights. Due to memory constraints, the entire 3D video is not actually queried at once; instead, each batch consists of the coordinates along a random set of projection rays at different times.

point is passed through the forward model and compared with measured projections, and the gradient of the loss function is backpropagated to update the network weights. There is no additional training data, and a new network is optimized from scratch for each new reconstruction.

The forward model we use to optimize our INR is parallel-ray linear projection: for each projection angle, projection along each ray is modeled as a sum of the values along that ray. This is a reasonable approximation to physical reality for some experiments, though some samples will exhibit nonlinear effects like multiple scattering. This simplified model is also motivated by constraints on GPU memory, since querying all 3D video coordinates at once is infeasible: each query requires saving a copy of the entire network for backpropagation. Instead, the queries need to be split into batches, but each batch must lead to a gradient update: it must generate a prediction that can be passed through the forward model and used to compute the loss function. The linear projection forward model makes it possible to compute the loss function using just the coordinates along a single ray through the 3D volume, since each projection ray determines the value of a measurement pixel. This forward model allows our method to not be memory-constrained by the size of the 3D video to be reconstructed, and can easily scale to large data volumes. In practice, each batch is a random set of rays across all time points, in order to avoid biasing the solution towards any particular time or spatial location.

3 Simulation and Results

To generate simulated carbon-contaminated phase projections, we created a simplified model of carbon accumulating on the object over time. We started with the atomic structure from [2] as our sample, which is a double-walled carbon nanotube encapsulating a sandwich structure of zirconium and tellurium. Next, we generated a layer of randomly placed carbon atoms on the sample surface, enforcing the minimum possible distance between atoms. Then, starting at $t = 0$ with no carbon, we gradually added this carbon layer, proceeding from one side of the sample to the other. Though this does not reflect how carbon accumulates in reality, it is sufficient as a test case for time-varying algorithms. abTEM [9] was used to generate the 3D atomic potential at each time point, which was fed into our linear forward model (Sec. 2) to generate simulated phase projections at each angle/time point. 181 projection angles/time points were used, with angles in $[0°, 180°]$ at $1°$ increments. These parameters were selected as idealized conditions, and future work extending this work to more realistic experimental conditions will need to account for sparser angular sampling ($\approx 3°$), a more restricted range of angles ($\approx 120°$), alignment issues, noise, and a larger number of voxels.

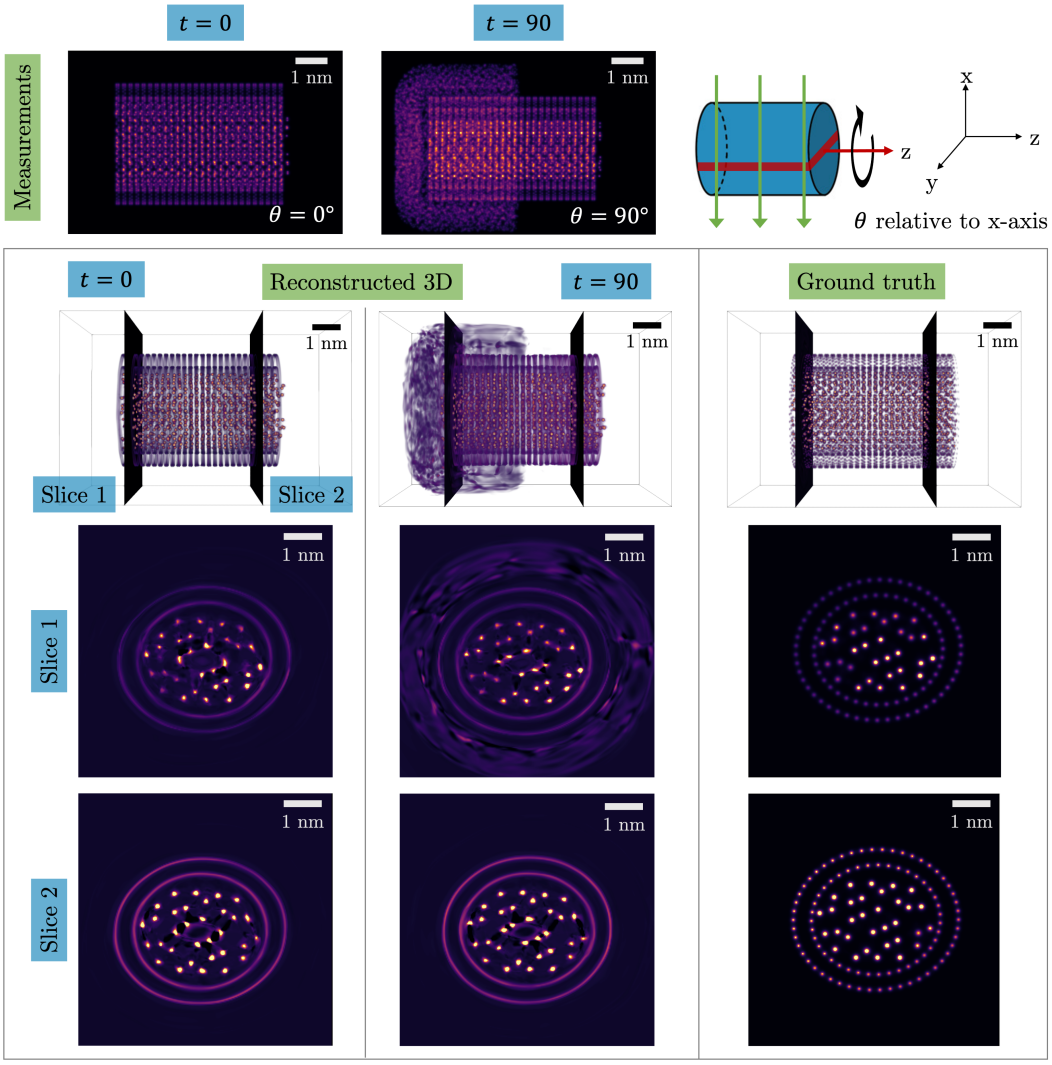

Figure 3: We generated simulated carbon-contaminated phase projections of a carbon nanotube sample, shown in the top row. The reconstruction from our space-time implicit neural representation (INR) is shown at time points $t = 0$ and $t = 90$ and compared to the uncontaminated ground truth. We see that at both time points, the heavier atoms in the inner structure of the tube are well resolved and the axial layers of the tube are well separated. These results demonstrate that even though the INR only directly sees one 2D projection at each time point, its implicit data prior successfully combined 3D information from different contaminated measurements.

Figure 3 shows preliminary results of using our space-time INR to reconstruct from these simulated carbon-contaminated measurements. Overall, the 3D reconstruction of the carbon nanotube's inner structure is accurate and reasonably sharp, with clear separation between layers of the structure and between atoms. Two 3D reconstructions at time points $t = 0$ and $t = 90$ can be used to show that the INR captured both the uncontaminated sample structure and the temporal dynamics of carbon contamination. The carbon atoms in the two outer walls of the carbon nanotube are not resolved by our reconstruction, but this appears to be unrelated to dynamic contamination since it is also present in an INR reconstruction of a static sample. It is more likely a challenge related to convergence of the INR optimization for high spatial frequencies.

In conclusion, we have demonstrated promising preliminary results on simulated data using a space-time INR as a flexible and computationally tractable space-time model for the carbon contamination problem in ptychographic atomic electron tomography (PAET). The relatively simple architecture of a

single MLP with no explicit regularization is nonetheless able to provide a high-quality reconstruction on simulated data on this highly underdetermined inverse problem. These results show that the implicit prior of the INR strongly influences the optimization process towards physically plausible solutions, without the help of explicit regularizers that are so crucial to traditional space-time methods. It is an important challenge for future work to further test the validity of this implicit prior, especially as these methods are increasingly applied to problems of scientific discovery.

References

[1] Colin Ophus. Four-Dimensional Scanning Transmission Electron Microscopy (4D-STEM): From Scanning Nanodiffraction to Ptychography and Beyond. *Microscopy and Microanalysis*, 25(3):563–582, June 2019. Publisher: Cambridge University Press.

[2] Philipp M. Pelz, Sinead Griffin, Scott Stonemeyer, Derek Popple, Hannah Devyldere, Peter Ercius, Alex Zettl, Mary C. Scott, and Colin Ophus. Solving Complex Nanostructures With Ptychographic Atomic Electron Tomography, June 2022. arXiv:2206.08958 [cond-mat, physics:physics].

[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):1–14, August 2018.

[4] Frank Ong, Xucheng Zhu, Joseph Y. Cheng, Kevin M. Johnson, Peder E. Z. Larson, Shreyas S. Vasanawala, and Michael Lustig. Extreme MRI: Large-scale volumetric dynamic imaging from continuous non-gated acquisitions. *Magn Reson Med*, 84(4):1763–1780, October 2020.

[5] Albert Pumarola, Enric Corona, Gerard Pons-Moll, and Francesc Moreno-Noguer. D-NeRF: Neural Radiance Fields for Dynamic Scenes. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10313–10322, June 2021. ISSN: 2575-7075.

[6] Keunhong Park, Utkarsh Sinha, Jonathan T. Barron, Sofien Bouaziz, Dan B Goldman, Steven M. Seitz, and Ricardo Martin-Brualla. Nerfies: Deformable Neural Radiance Fields. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5845–5854, Montreal, QC, Canada, October 2021. IEEE.

[7] Ruiming Cao, Fanglin Linda Liu, Li-Hao Yeh, and Laura Waller. Dynamic Structured Illumination Microscopy with a Neural Space-time Model. Technical Report arXiv:2206.01397, arXiv, June 2022. arXiv:2206.01397 [physics] type: article.

[8] Matthew Tancik, Pratul P. Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan T. Barron, and Ren Ng. Fourier Features Let Networks Learn High Frequency Functions in Low Dimensional Domains. *arXiv:2006.10739 [cs]*, June 2020. arXiv: 2006.10739.

[9] Jacob Madsen and Toma Susi. abTEM: ab Initio Transmission Electron Microscopy Image Simulation. *Microscopy and Microanalysis*, 26(S2):448–450, August 2020. Publisher: Cambridge University Press.