Electrostatic Potential prediction using a SwinUNETR

Authors (alphabetical): Andrea Cicconardi, Mattia Lizzano, Sridurgesh Ravichandran, Haipei Shao, Zeyu Wang

1. Introduction

Integrated Differential Phase Contrast (iDPC) enables live visualization of phase changes and has proven particularly powerful for imaging light elements. However, iDPC relies on several strong physical assumptions, including the weak-phase object approximation and linear imaging conditions. These assumptions progressively break down for thicker samples, where multiple scattering and electron channelling significantly modify the probe-sample interaction. Also, iDPC is highly sensitive to residual specimen mis-tilt, which introduces systematic phase gradients and distortions that can obscure the true projected electrostatic potential[1].

In this project, we explore a data-driven alternative to conventional iDPC reconstruction. Instead of analytically inverting segmented detector signals, we train a SwinUNETR to directly map four-segment annular dark-field (ADF) images to the underlying electrostatic potential, using physically realistic simulations as ground truth.

2. Physical Limitations of iDPC

For a thin specimen we can assume that the phase shifts are small, and we can neglect multiple scattering. Under this weak phase object approximation (WPOA), the exit electron wave can be written as:

$$\varphi(r) \approx e^{-i\sigma V(r)} \approx 1 + i\sigma V(r)$$

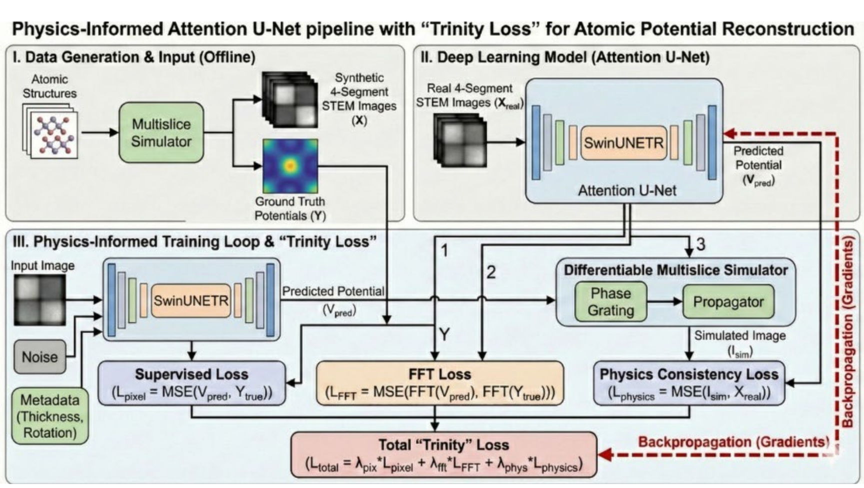
Where V(r) represents the projected electrostatic potential and σ the electron-matter interaction constant. The iDPC method reconstructs the projected potential by integrating the DPC signals in reciprocal space:

$$V(k) \propto \frac{k_x D_x(k) + k_y D_y(k)}{i|k|^2 (H)_k}$$

For thicker samples, iDPC background becomes messy and reconstruction of V(r) is no longer reliable.

3. Workflow

<u>*Physics-constrained Data Generation*</u>: We constructed a metadata-rich synthetic dataset by coupling first-principles Density Functional Theory (DFT) with multislice electron scattering. Ground-truth electrostatic potentials were computed using CP2K for five structurally diverse crystalline systems (e.g., $KNbO_3$, GaN, perovskite), preserving full atomic resolution. These potentials were used to drive high-resolution STEM simulations in abTEM, generating four-segment ADF detector images at 0.1 Å sampling (256×256 pixels). To induce realistic dynamical scattering artifacts, simulations spanned wide thickness ranges (up to 150 nm) and randomized crystal mis-tilts, well beyond the weak-phase regime. Importantly, exact physical parameters—including thickness and Euler angles—were retained as spatially broadcast metadata, enabling conditional learning of the non-linear relationship between scattering-induced contrast distortions and the underlying electrostatic potential.



Powerful Inversion Model: We employ a SwinUNETR architecture, which combines a hierarchical Swin Transformer encoder with a U-Net style decoder[3]. The transformer encoder captures non-local correlations through window-based self-attention, which is particularly relevant for modelling multiple scattering effects in thick samples. The decoder, together with skip connections, enables accurate pixel-wise reconstruction of the potential.

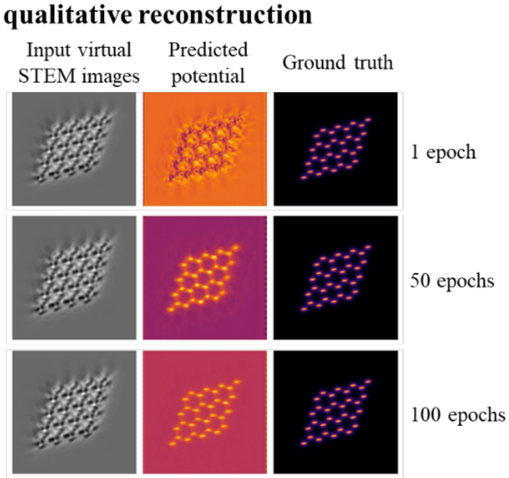
Model Running: The learning task is formulated as an inverse problem. Given four segmented ADF images acquired under conditions where analytical inversion fails, we train the architecture to predict the corresponding projected potential. The training objective minimizes the discrepancy between the predicted projected potential and ground truth from DFT calculation. Training is performed using a physics-inspired loss function that combines a pixel-wise L1 loss with an additional frequency-domain loss computed on the Fourier magnitude of the predicted and target potentials. While the pixel loss enforces local accuracy, the frequency loss encourages correct spatial frequency content, which is essential for preserving physically meaningful contrast and periodicity.

4. Results

We present qualitative reconstruction results after 1, 50, and 100 training epochs (Fig. **a**), training convergence (Fig. **b**) and summarize the corresponding quantitative performance metrics (Fig. **c**). Reconstructions obtained using conventional iDPC are shown as a reference for direct comparison under identical conditions (Fig. **d**). The robustness of the proposed method against specimen mis-tilt is further demonstrated (Fig. **e**), where mis-tilt induced background gradients and distortions are effectively suppressed.

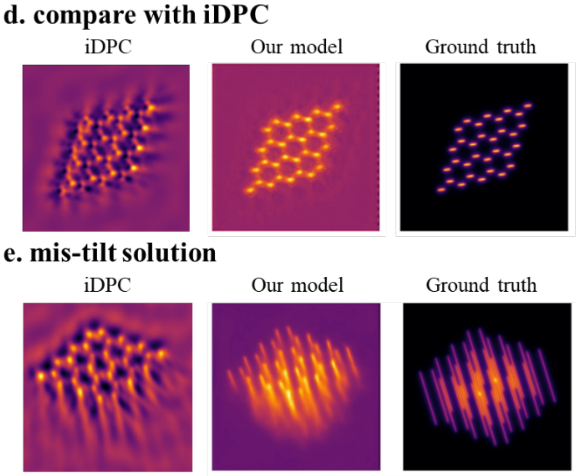
Quantitative validation of the reconstructed electrostatic potential is provided in Fig. **f**. Parity analysis comparing the predicted maximum electrostatic potential with the DFT ground truth for the validation set reveals a strong linear correlation ($R^2 > 0.9$), demonstrating that the model reliably recovers quantitative physical values (in volts) over a wide dynamic range. The isolated deviation observed at potentials exceeding > 350 kV corresponds to optically thick specimens, where strongly non-linear dynamical scattering effects dominate.

a. qualitative reconstruction

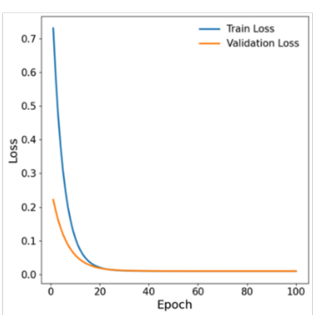


d. compare with iDPC

e. mis-tilt solution



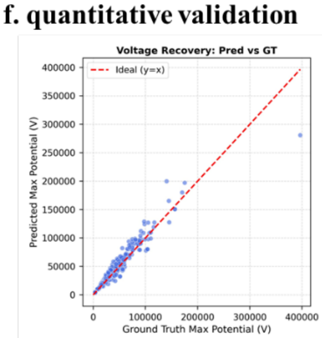
b. training convergence



c. quantitative performance metrics

Category	Metric	Value
Normalized Potential	MAE	0.001496
	RMSE	0.003071
	PSNR (dB)	54.456
	Pearson Correlation	0.7205
	Peak Error @ 99.5% (%)	23.82
Absolute Potential (V)	MAE (V)	897.84
	RMSE (V)	1842.38
	Peak Error @ 99.5% (%)	23.82
Potential Statistics	Ground Truth Max (V)	25,939.86
	Predicted Max (V)	28,327.29
	Ground Truth Mean (V)	2,162.52
	Predicted Mean (V)	1,825.91

f. quantitative validation



5. Key findings

The proposed approach can learn a physically meaningful mapping from segmented ADF images to electrostatic potential, even in regimes dominated by multiple scattering and mis-tilt:

(1) The trained model successfully enables a thickness-robust reconstruction strategy that remains effective far beyond the weak-phase assumption and explicitly mitigates mis-tilt induced artifacts.

(2) Large-scale, physics-constrained dataset provide high-fidelity and physically consistent training supervision.

(3) Powerful inversion model, based on a SwinUNETR transformer-augmented encoder–decoder architecture.

6. Perspectives

(1) The present model is trained at a fixed spatial resolution of 256×256 pixels, which constrains the achievable reconstruction fidelity. This limitation is not fundamental and can be systematically addressed by expanding the training dataset to higher resolutions and larger simulation volumes, thereby further improving spatial detail and generalization. (2) Extending the framework to experimental validation is a natural next step. Wedge-shaped specimens with thicknesses of 60–200 nm (FIB-prepared YSO sample) provide an ideal test platform, where thickness can be independently quantified by EELS and corresponding four-segment STEM images acquired across the wedge. Incorporating such thickness-calibrated experimental data will further enhance the robustness and practical applicability of the proposed inversion model.

Find full work at:

https://github.com/sridurgesh007/microscopy_hackathon_team_iphasenet.git

Ali Hatamizadeh, V. N. (2022). Swin UNETR: Swin Transformers for Semantic Segmentation of Brain Tumors in MRI Images. *arXiv:2201.01266*.

Eric GT Bosch, I. L. (2016). Integrated Differential Phase Contrast (iDPC) STEM: A New Atomic Resolution STEM Technique To Image All Elements Across the Periodic Table. *Microscopy and Microanalysis, 22*, 306-307. doi:https://doi.org/10.1017/S1431927616002385

Thomas D. Kühne, e. a. (2020, May 21). CP2K: An electronic structure and molecular dynamics software package - Quickstep: Efficient and accurate electronic structu re calculations. *J. Chem. Phys., 152 (19)*. doi:https://doi.org/10.1063/5.0007045