OPTICA

Accelerated deep self-supervised ptycho-laminography for three-dimensional nanoscale imaging of integrated circuits

Iksung Kang,[1,2] Yi Jiang,[3] Mirko Holler,[4] Manuel Guizar-Sicairos,[4,5] A. F. J. Levi,[6] Jeffrey Klug,[3] Stefan Vogt,[3,9] and George Barbastathis[7,8,10]

[1]*Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[2]*Current address: Department of Molecular and Cell Biology, University of California, Berkeley, California 94720, USA*
[3]*Argonne National Laboratory, Lemont, Illinois 60439, USA*
[4]*Paul Scherrer Institut, Forschungsstrasse 111, 5232 Villigen PSI, Switzerland*
[5]*Institute of Physics (IPHYS), Ecole Polytechnique Fédérale de Lausanne, Rte Cantonale, Lausanne 1015, Switzerland*
[6]*Department of Electrical and Computer Engineering, University of Southern California, Los Angeles, California 90007, USA*
[7]*Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*
[8]*Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 Create Way, Singapore 117543, Singapore*
[9]*svogt@anl.gov*
[10]*gbarb@mit.edu*

Three-dimensional inspection of nanostructures such as integrated circuits is important for security and reliability assurance. Two scanning operations are required: ptychographic to recover the complex transmissivity of the specimen, and rotation of the specimen to acquire multiple projections covering the 3D spatial frequency domain. Two types of rotational scanning are possible: tomographic and laminographic. For flat, extended samples, for which the full 180° coverage is not possible, the latter is preferable because it provides better coverage of the 3D spatial frequency domain compared to limited-angle tomography. It is also because the amount of attenuation through the sample is approximately the same for all projections. However, both techniques are time consuming because of extensive acquisition and computation time. Here, we demonstrate the acceleration of ptycho-laminographic reconstruction of integrated circuits with 16 times fewer angular samples and 4.67 times faster computation by using a physics-regularized deep self-supervised learning architecture. We check the fidelity of our reconstruction against a densely sampled reconstruction that uses full scanning and no learning. As already reported elsewhere [Opt. Express 28, 12872 (2020)], we observe improvement of reconstruction quality even over the densely sampled reconstruction, due to the ability of the self-supervised learning kernel to fill the missing cone. © 2023 Optica Publishing Group under the terms of the Optica Open Access Publishing Agreement

https://doi.org/10.1364/OPTICA.492666

1. INTRODUCTION

Hard x rays offer non-destructive visualization and metrology of nanoscopic details inside complex structures, such as integrated circuits (ICs). Short wavelength and long penetration depth of x rays enable probing into the volumetric interiors of ICs. X-ray imaging instruments often incorporate the object rotation with respect to x-ray illumination to improve the depth resolution of reconstructions. It is desirable to select the rotation geometry by carefully considering the objects' geometrical properties. For instance, for flat extended nanostructures such as ICs, the object's rotation axis can be oblique to the direction of synchrotron x rays, i.e., laminographic imaging [1,2]. For flat, extended samples, the oblique geometry keeps the amount of x-ray absorption and scattering by the structures approximately the same regardless of

the object rotation, so that the volumetric interiors of ICs can be more reliably reconstructed. This makes a clear distinction from existing tomographic imaging methods [3–6], where the strengths of absorption would vary across different rotation angles.

On the other hand, translational scanning of the object enables a larger field of view reconstruction. Originally proposed for scanning transmission electron microscopy (STEM), ptychography leverages lateral movement of either the object or the illumination to acquire several diffraction patterns from different lateral locations to add robustness to phase retrieval [7–10] and to reconstruct a larger field-of-view object [11–13]. Objects are computationally retrieved from the ptychographic measurements by some well-established algorithms, such as a ptychographic iterative engine (PIE) [11], difference map (DM) [14], least-square maximum likelihood (LSQ-ML) [15], etc. Alternatively, ptychography may also

be conducted in the Fourier domain by replacing the illumination with a set of plane waves incident at different angles [16–18].

As ICs have a flat geometry and an extended field of view, ptychography becomes synergistic with laminography in imaging such nanostructures. Holler *et al.* demonstrated ptychographic x-ray laminography on ICs fabricated with 16-nm technology with 18.9-nm resolution [19] using the laminographic nano-imaging instrument (LamNI) [20], whose reconstruction quality easily

surpassed that of x-ray ptychographic tomography [6]. In a typical x-ray ptycho-laminographic imaging apparatus as depicted in Fig. 1(A), synchrotron x rays obliquely illuminate ICs with the angle of θ between the direction of x-ray propagation and the rotation axis. ICs are scanned from a few thousand laminographic angular views, where angular sampling depends on the sample thickness, the resolution, and the laminographic angle [19], and
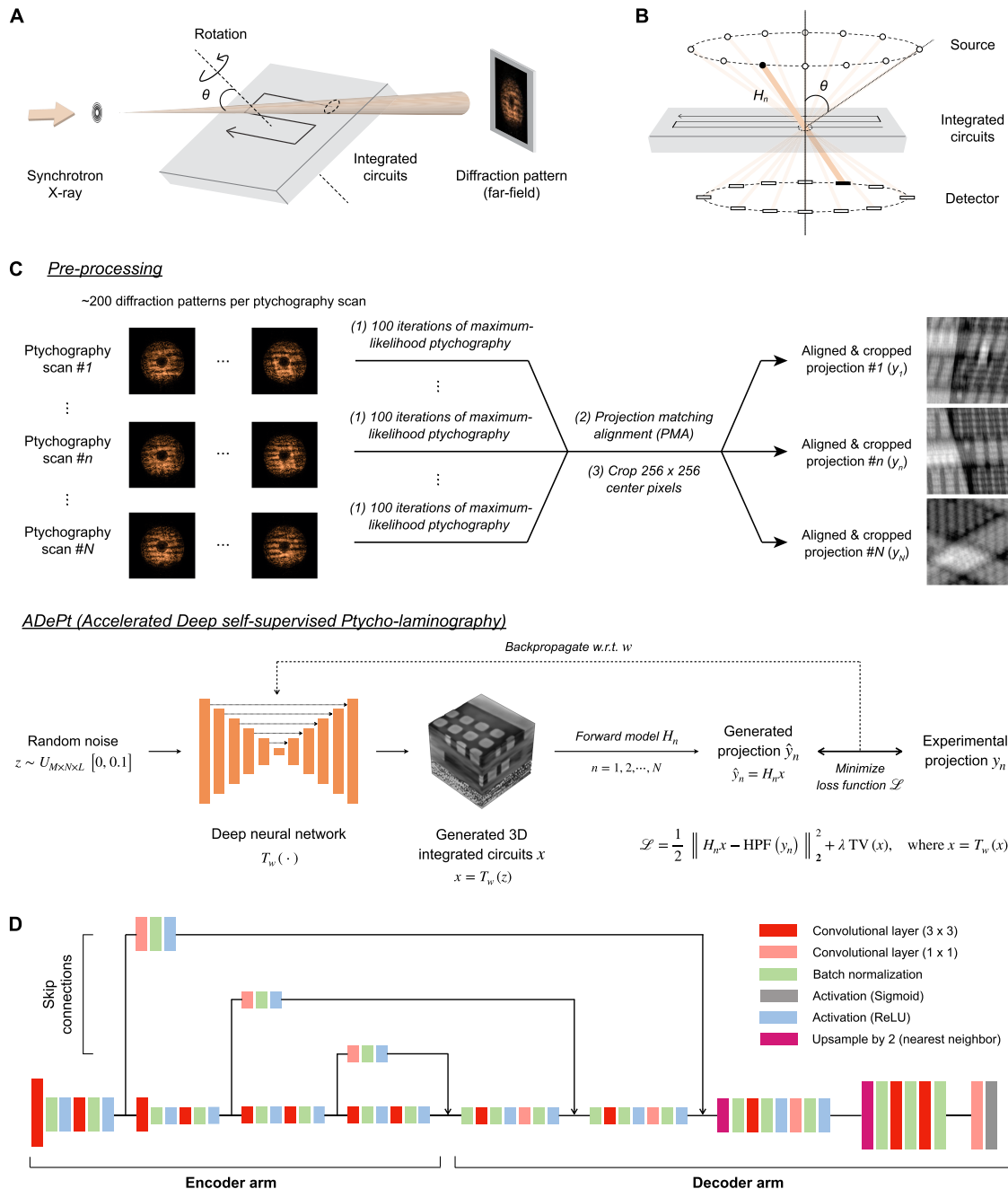
Fig. 1. Accelerated deep self-supervised ptycho-laminography. (A) Ptycho-laminographic imaging geometry. Synchrotron x rays illuminate a sample of integrated circuits in the ptycho-laminography geometry, with the sample rotating around the oblique laminographic axis and scanned over a few thousand angles. For each ptychography scan, the sample is laterally scanned at several hundred different locations. (B) Equivalent imaging geometry. Forward operators H_n ($n = 1, 2, \ldots, N$) are defined according to each laminographic rotation. (C) Proposed physics-informed machine learning framework. Our pre-processor translates experimental ptycho-laminographic measurement from the detector plane to the sample domain with minimal processing using a ptychographic reconstruction algorithm. ADePt generates a 3D image of integrated circuit morphology from the pre-processed projections throughout the optimization process. (D) Deep neural network architecture for self-supervised learning. The proposed architecture is essentially an encoder–decoder convolutional neural network with skip connections and receiving random noise as input. The output is the image. Code is publicly available at [21].

for each ptychography scan, x rays laterally scan the ICs at several hundred different lateral positions.

Three-dimensional information of nanostructures is retrieved using a two-step iterative update process based on our ptycho-laminographic measurements, i.e., reconstruction from the densely sampled dataset (or *densely sampled reconstruction*): (1) a thousand iterations of a ptychographic reconstruction algorithm processes hundreds of diffraction patterns to get a projection for each of the 2000 ptychograms (more details about the reconstruction are in Section 2.B); and (2) a volumetric reconstruction is formed by the laminographic synthesis of all projections. The missing cone in the Fourier domain is filled in during the post-processing step. The densely sampled reconstruction, however, results in a long data acquisition and computation time due to the strict angular sampling requirement [19] of ptychograms and the iterative reconstruction of projections.

Here, we demonstrate up to an aggregate 9.57-fold time savings in x-ray ptycho-laminographic reconstruction by using a physics-regularized deep self-supervised learning architecture, called accelerated deep self-supervised ptycho-laminography (ADePt). We achieved this acceleration by reducing the number of angular samples by 16 times and the computation time by 4.67 times. Considering the memory limitations in GPUs, ADePt yields the reconstruction of $4.36 \times 4.36 \times 3.92\text{-}\mu m^3$ ICs within 2 h; see Table S1 for more timing details. We provide both quantitative and qualitative comparisons on the performance of ADePt with the densely sampled reconstruction using the bit-error ratio (BER) and 3D power spectral density. Finally, we observe that the self-supervised learning kernel fills missing cones from a much fewer number of projections compared to the densely sampled reconstruction.

ADePt consists of two components: a lightweight pre-processor and a deep neural network as illustrated in Fig. 1(C). The pre-processor consists of a few iterations of maximum likelihood ptychography, projection matching alignment (PMA), and center cropping, and works as an approximate inverse operator on diffraction patterns. The network parameterizes a 3D structure of ICs by considering both network-structure and physics priors. The proposed learning approach makes significant advances from the baseline in terms of the following: (1) ADePt does not use the ground truth structure of ICs for its learning process, which sets it apart from any supervised machine learning approaches and saves a significant amount of time and resources from the ground truth preparation, and operates on only a single set of experimental measurements; and (2) ADePt explicitly leverages the physical forward model of the x-ray ptycho-laminography geometry, which tightly regularizes the solution space and guides the algorithm to generate a physically feasible reconstruction even when the number of projections is largely reduced.

2. METHODS

A. X-Ray Ptycho-Laminography Experiment

Ptycho-laminography measurements were carried out at the coherent small-angle x-ray scattering (cSAXS) beamline at the Swiss Light Source at the Paul Scherrer Institut (PSI), Switzerland. An IC produced with 16-nm technology was scanned with synchrotron x rays of 6.2 keV using LamNI [20]. The laminographic angle between the rotation axis and the beam propagation axis was fixed to be 61°, and the angular step of projections was 0.18°, making 2000 scans in total. Diffraction patterns were recorded by a step scan at around 200 lateral locations with an in-vacuum Eiger 1.5-M detector (pixel size: 75 μm, sample-to-detector distance: 5.23 m, exposure time: 0.1 s) [22].

B. Baseline Method and Densely Sampled Reconstruction Preparation

Densely sampled reconstruction preparation takes experimental ptycho-laminographic measurements from 2000 ptychography scans and uses a three-step baseline method: (1) for every ptychography scan, projections are retrieved by 1000 iterations of the least-square likelihood ptychographic algorithm [15] as implemented in PtychoShelves [23]. This ptychographic reconstruction step takes far-field x-ray diffraction patterns (512×512 px^2, pixel size: 75 μm) recorded at 200 different lateral locations. (2) The projections are precisely aligned to each other using the PMA algorithm [24]. (3) The aligned projections are synthesized to reconstruct a volumetric structure using the standard Fourier backprojection (FBP) method, followed by the recovery of missing cone information. Data processing exactly follows the steps in [19]. These steps result in the densely sampled reconstruction with the voxel size of 27.2 nm, which is larger than the resolution of 19 nm reported in [19] due to a fewer number of projections acquired and a larger step size for each scan.

Despite the intensive densely sampled reconstruction, it remains ambiguous especially for longitudinal layer features, as a missing cone exists in the k-space due to the oblique x-ray illumination in the ptycho-laminographic imaging [19].

C. Architecture

The network design is based on a modified U-net architecture with a random input noise, following the implementation of a deep image prior [25], which has been widely used for many computational imaging applications, including image dehazing [26], super-resolution [25,27], phase retrieval [28,29], tomography [30,31], and magnetic resonance imaging (MRI) [32]. This implementation shares a similarity with coordinate-based learning approaches [33–36]. The deep neural network is implemented as an encoder–decoder architecture with skip connections [37] as shown in Fig. 1(D). Network weights are initialized with Xavier uniform distribution [38] (gain: 0.2). Following the convention of a deep image prior [25], random input noise z is given to the network, sampled from a uniform distribution $z \sim U_{M \times N \times L}[0, 0.1]$. The encoder reduces the lateral dimensions by a factor of four, and the decoder restores the dimensions back to the original. The architecture is chosen to have skip connections to relay the encoder features to the decoder according to our finding that an hourglass architecture does not reliably render high-frequency details. Finally, a sigmoid-like activation function sets the range of output values of the ICs within $[-0.03, 0.03]$.

D. Pre-processing and Network Optimization

Figure 1(C) shows our pre-processor in three steps: (1) 10 times fewer iterations, i.e., 100, of the LSQ-ML ptychographic algorithm are run to obtain intermediate projections; (2) the PMA algorithm aligns the intermediate projections to match with each other, which is applied to both densely sampled and reduced datasets; and (3) 256×256 center pixels are cropped from the

aligned projections to be used for the reconstruction process with ADePt. Figures S1 and S2 in Supplement 1 show a qualitative and quantitative analysis, summarizing the sensitivity to the alignment of projections.

ADePt iteratively updates the randomly initialized weights in the deep neural network according to the loss functional

$$\hat{w} = \text{argmin}_w \left[\frac{1}{2} \sum_{n=1}^{N} \| H_n x - \text{HPF}(y_n) \|_2^2 + \lambda \text{TV}(x) \right],$$

$$\text{where} \quad x = T_w(z) .$$

(1)

$T_w(\cdot)$ is a deep neural network that takes random noise z as input, where w represents its parameters, H_n a physical forward model corresponding to the nth ptychography scan, y_n a pre-processed projection of the nth ptychography scan, HPF$(\cdot)$ a high-pass filter on y_n's, and TV$(\cdot)$ a total-variation (TV) regularization operator acting upon the 3D IC structure x.

The problem of determining the structure x from given diffraction patterns is highly ill posed, primarily because of three factors: (1) a tenfold reduction in the number of maximum-likelihood ptychography iterations to acquire projections, (2) a decrease of eight to 16 times in the number of projections used for the reconstruction, and (3) the general tendency of low spatial frequencies to dominate the training of neural networks as we have pointed out before [39]. To address the imbalanced influence of the low spatial frequencies, we enforced a stronger prior during network training by applying a high-pass filter to the projections, following a similar approach as described in [39]. To achieve this, we used wavelet decomposition to design the filter. The projections were decomposed into four levels of wavelet scattering coefficients using the Daubechies wavelet (db8) [40]. This facilitates the clear identification of the low- and high-spatial-frequency components through visualization at different levels, making it a better alternative to a fast Fourier transform (FFT)-based high-pass filter. For each subsequent level of decomposition, the wavelet coefficients were multiplied by a factor of 2.5. As a result, this method effectively amplifies the emphasis on higher spatial frequencies in the projections while ensuring that the total energy of the filtered projections remains normalized to that of the original projections.

While this approach enabled us to obtain a solution with accentuated fine details, it does come at the cost of unwanted high-spatial-frequency artifacts in the reconstruction (as depicted in Fig. 5(C)). To mitigate the unwanted artifacts, we incorporated a TV regularization term into the training loss function, which promotes sparsity in the image gradient and effectively suppresses the artifacts. Considering that the nominal spatial frequency of features in ICs varies with depth, the TV regularization parameter λ is set differently across the z axis to control the regularizer strength, i.e., $\lambda = 3 \times 10^{-6}$ for $z < 2.75$ µm and $\lambda = 3 \times 10^{-8}$ otherwise. We assigned the larger value to shallow layers with coarser features and the smaller value to deeper layers with finer features. The determination of coarse and fine features was based on the power spectral density of layer features in the densely sampled reconstruction, serving as the ground truth (refer to Fig. S3 in Supplement 1 for more details). Alternatively, one can refer to the circuit diagram used for fabrication to identify the layers with coarse and fine features and determine the appropriate range and values of the regularization parameter.

The reconstruction process is run for 1500 iterations using the Adam optimizer [41] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and the initial learning rate of 2×10^{-4}, which is halved after 1000 iterations.

3. RESULTS

A. Qualitative Performance Comparison

We prepare the densely sampled reconstruction for the performance comparison using the two-step iterative method with 2000 projections sampled every 0.18° angular increment $\Delta\varphi$. Ptycho-laminographic measurements are processed with 1000 iterations of the LSQ-ML algorithm for each ptychography scan, followed by the PMA, the filtered backprojection (FBP), and the recovery of missing cone information [19].

In this study, we assess ADePt's 3D physically feasible rendering performance under sparsely sampled conditions by making a comparison between the densely sampled reconstruction and the ADePt reconstructions. Here, we increase $\Delta\varphi$ from 0.18° to 1.44° and 2.88°, limiting the number of available projections to 250 and 125 from 2000. In Fig. 2, we visualize reconstructed IC features at five different depths, i.e., z-axis locations, using the two-step baseline and ADePt with 250 and 125 projections to qualitatively compare them with the densely sampled reconstruction. Visually comparing the two, ADePt reconstructions are more closely aligned to the densely sampled reconstruction for both cases—even better than the densely sampled reconstruction with greater feature contrast to identify high-spatial-frequency details that come at the cost of the ability to quantify—although the FBP reconstructions show much lower spatial resolution due to accrued artifacts from the missing cones and sparse sampling. We support the argument with the aid of the cross-section profiles. Figure 3 visualizes the reconstructions from different axial views. Moreover, ADePt benefits from the sparse sampling scheme, resulting in ×6.58 and ×9.57 aggregate reduction in computation time with 250 and 125 projections, respectively, compared to the densely sampled reconstruction. For more details on time breakdown, please see Table 1.

B. Power Spectral Density Representation

It is easier to describe the aforementioned artifacts in a different domain as both the terms sparse sampling and missing cones are defined based on the k-space. We visualize both the densely sampled reconstruction and FBP and ADePt reconstructions as k-space representations by means of power spectral density, which is the Fourier transform of the autocorrelation function, i.e., the Wiener–Khinchin theorem.

In Fig. 4, we provide 2D power spectral density profiles of the baseline and ADePt reconstructions using 250 and 125 projections along with the densely sampled reconstruction. Both FBP reconstructions show their missing cones rooted in an oblique rotation axis in the ptycho-laminography geometry. Sparse-sampling artifacts are also easily noticed in the reconstructions, which provides explanations on lack of spatial resolution as shown in Fig. 2.

Our framework, however, learns from experimental ptycho-laminographic measurements how the missing parts should be filled to represent physically feasible IC solutions guided by the physical and network-structure priors. Moreover, comparing the densely sampled reconstruction and ADePt reconstructions in k-space, we notice that ADePt renders ICs even better—the
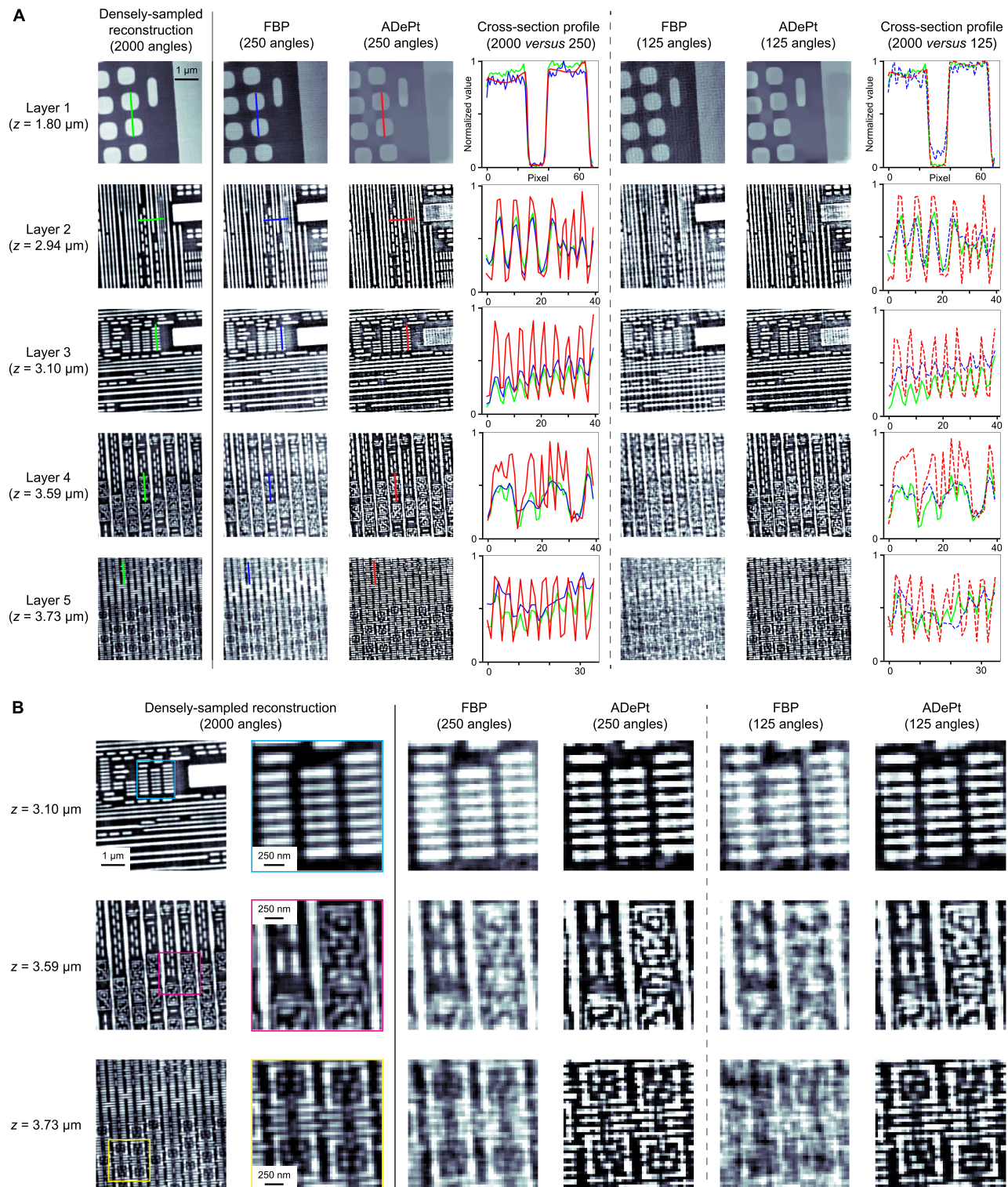
Fig. 2. Comparison between FBP and ADePt's reconstructions and the densely sampled reconstruction. (A) We exploit physics-informed machine learning to reliably reconstruct integrated circuits with a reduced number of projections, i.e., 125 and 250 out of 2000, and qualitatively compare the FBP and ADePt reconstructions with the densely sampled reconstruction. Please note that the colormaps used in some reconstructions are selected differently. Specifically, for (1) the densely sampled reconstruction, the colormaps of layers 2–5 are set to its 2.5th and 80th percentiles, and for (2) the FBP reconstructions with both 125 and 250 angles, the colormaps of layers 2–5 are fixed to its 12.5th and 80th percentiles. For all other figures, the colormaps are fixed to the minimum and maximum values of their respective reconstructions. (B) We qualitatively compare the reconstructions within their respective zoomed-in areas for better evaluation. The colormap conventions followed in (A) are also applied.
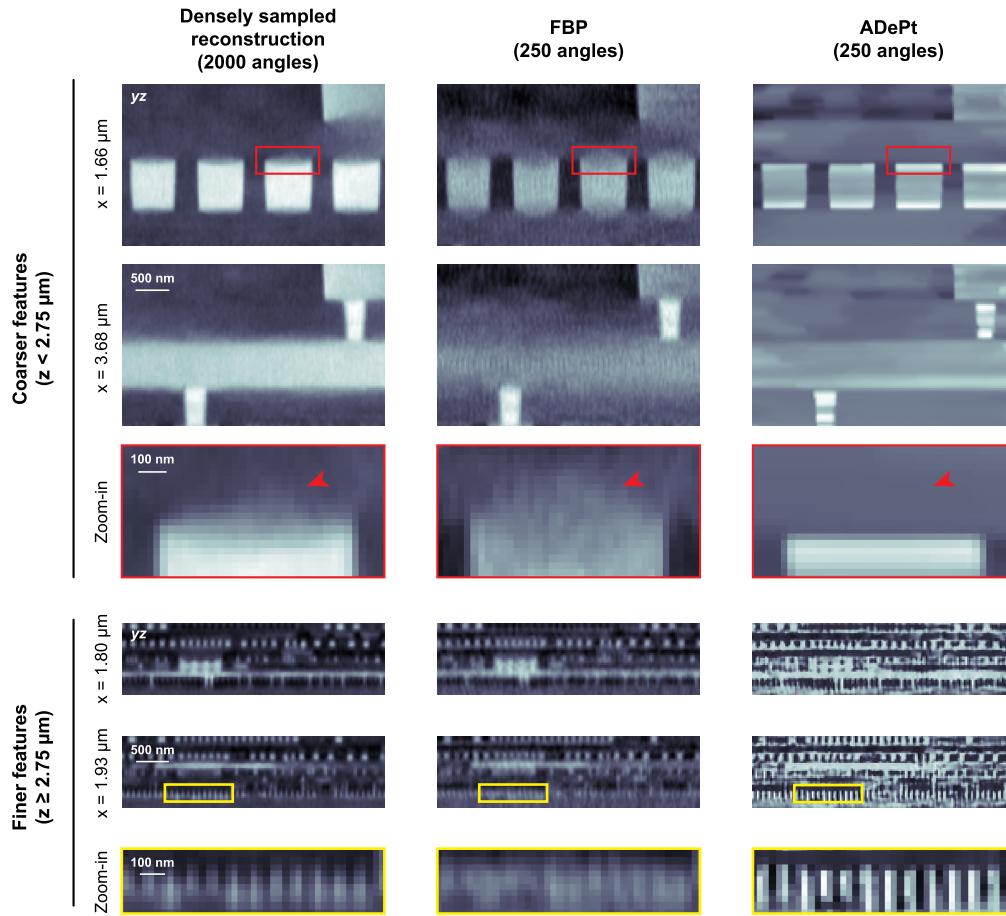
Fig. 3. Qualitative comparison among yz cross sections of different reconstructions. Although the densely sampled reconstruction yields the best contrast among coarser features, missing cone artifacts are not completely addressed (see red arrows). Some finer features can be displayed more clearly with ADePt than with the FBP reconstructions and the densely sampled reconstruction, at the cost of ability to quantify (see yellow boxes).

Table 1.   Data Preparation and Computation Time Breakdown

	Densely Sampled Reconstruction (2000 Angles)	ADePt (250 Angles)	ADePt (125 Angles)
Data acquisition[a]	11.1	1.38 (×8 faster)	0.694 (×16 faster)
Data computation[b]	4.27	0.958 (×4.46 faster)	0.914 (×4.67 faster)
(1) Iterative update (maximum likelihood)	4.05[c]	0.0890	0.0445
(2) Reconstruction - data I/O, alignment, synthesis	0.218[d]	0.869	0.869
Total time (h)	15.4	2.34 (×6.58 faster)	1.61 (×9.57 faster)

[a]Data acquisition uses LamNI [20] to acquire ptycho-laminographic measurements over the integrated circuit sample ($26.2 \times 38.2 \times 3.92$ μm³).

[b]Data computation includes an iterative ptychographic update step, data I/O, projection matching alignment [24], and laminographic synthesis. Densely sampled reconstruction uses 1000 iterations of the LSQ-ML algorithm [15,23], and ADePt reconstructions use 100 iterations of the same algorithm. The data computation time breakdown comparison is made on the reconstructed sample ($4.36 \times 4.36 \times 3.92$ μm³).

[c]Iterative ptychographic update of the densely sampled reconstruction is performed on 10 GTX 1080 GPUs.

[d]Data reconstruction process of the densely sampled reconstruction is based on 1 V100 GPU. All other computations use 2 V100 GPUs.
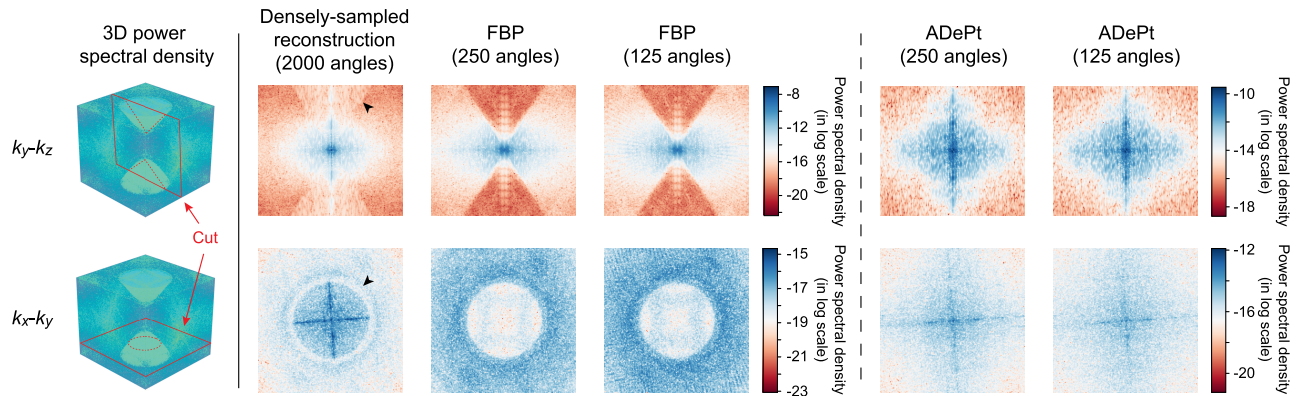
missing cone of the densely sampled reconstruction remain unintentionally accentuated, as shown with black arrows in Fig. 4, whereas the cone filled and spectrum recovered with ADePt is more continuous. As already reported elsewhere [42], it is plausible that ADePt could outperform the baseline method even with more limited amount of data given as the proposed framework takes a self-supervised approach, thus completely agnostic to the type of specimen. However, this does not preclude future use of priors to boost performance.

C. Quantitative Evaluation

We use BER as a metric to quantify the ratio of erroneous occupancy in the reconstructions with reference to the densely sampled

Fig. 4.   Power spectral density analysis for qualitative comparison. We visualize FBP and ADePt reconstructions and the densely sampled reconstruction in k-space to visualize artifacts due to missing cone and sparse sampling. Cuts are made along the $k_y - k_z$ and $k_x - k_y$ planes (red arrows). We demonstrate that ADePt provides reconstructions with fewer artifacts in k-space, considering that the FBP reconstructions display artifacts due to the angular subsampling and missing cone, and that the densely sampled reconstruction shows artifacts due to imperfect missing cone filling (black arrows).

reconstruction as the ground truth [43]. Binary objects such as ICs are particularly suited to this metric for quality assessment. Although many printing materials comprise ICs, including copper, tungsten, and aluminum, here we just treat the ICs as binary with regard to the occupancy, irrespective of material. Using the expectation–maximization (EM) algorithm in the context of Gaussian mixture models, we first binarize the densely sampled reconstruction to get the ground truth. However, since the densely sampled reconstruction may still be ambiguous especially for longitudinal features due to missing cones in the Fourier domain (see Fig. 4), the layers with ambiguous features will not be accurately binarized. Thus, we exclude the layers with binarization errors from our quantitative analyses in Section 3.C. More details can be found in Fig. S5 in Supplement 1.

In Fig. 5(A), we compare ADePt's performance with the baseline method's using BER and the Pearson correlation coefficient (PCC). We demonstrate that ADePt outperforms the baseline method for both conditions with fewer projections, but the difference between the two methods becomes more statistically significant when only 125 projections are considered during the reconstruction process, which suggests that ADePt is more tolerant to sparse sampling. The layers located at $z > 2.75$ μm are considered to have finer circuit features, according to typical design rules of ICs (Fig. S3 in Supplement 1 illustrates this further). This implies that higher-frequency details are better reconstructed with the physics-informed self-supervised machine learning, even in regions with fine transverse details.

D. Ablation Study

As illustrated in the loss function in Fig. 1(C), we employ two key design elements to the implementation: (1) high-pass filtering on the intermediate projections to improve feature contrast and to enforce high-frequency bias to the network; and (2) TV regularization in the loss functional to suppress residual artifacts. We assess the relative contribution of each design element in our proposed framework by ablating each one of the following elements in succession: high-pass filter on projections and TV regularization.

The results are shown in detail in Fig. 5. One salient observation is that in the region of all features (z between 0 and 3.92 μm), performance degrades by equal amounts if the high-pass filter or

TV term is ablated. On the other hand, in the fine feature regime (z between 2.75 and 3.92 μm), the significance of the high-pass filter is higher, as evidenced in Fig. 5(B) by the catastrophic drop in performance when it is ablated. Visual inspection of the ablated reconstructions in Fig. 5(C) confirms these trends.

4. DISCUSSION

ADePt provides a 3D estimate of ICs from experimental ptycho-laminographic measurements using deep self-supervised learning. The proposed framework makes explicit use of a physical forward model to obtain the image iteratively, and regularizes through the chosen sparsity-enforcing neural network kernel and an additional TV penalty in the loss function. This is a significant departure from supervised learning approaches that typically require ground truth to prepare a paired dataset for training, which is expensive.

Supervised methods' performance is also limited in the absence of accurate ground truth. For IC imaging, this is almost always the case since obtaining the true shape is challenging. In earlier supervised work [43], we used reconstructions from dense angular sampling as ground truth. Here, we observe that ADePt fills the missing cone more effectively than the densely sampled reconstructions, as Fig. 4 clearly illustrates. That the IC geometry is very compatible with our chosen auxiliary TV regularizer strengthens this claim. Further investigation of the ADePt scheme's performance in different types of specimens is a good topic for future work.

In this work, we treat all ptycho-laminographic projections as equal. However, in previous work [44], we have demonstrated that there is a benefit to weighing each differently, according to an attentional scheme [45,46]. Also, coordinate-based learning methods [33–36] may be beneficial to further increase the reconstruction volume, as the methods generally take fewer trainable parameters than convolutional neural network architectures. Finally, as an alternative to using the high-pass filter and TV regularization, one could consider employing nonlinear diffusion regularization [47] to handle slowly spatially varying backgrounds, as depicted in "(-) HPF" in Fig. 5(C). We leave this for future work as well.
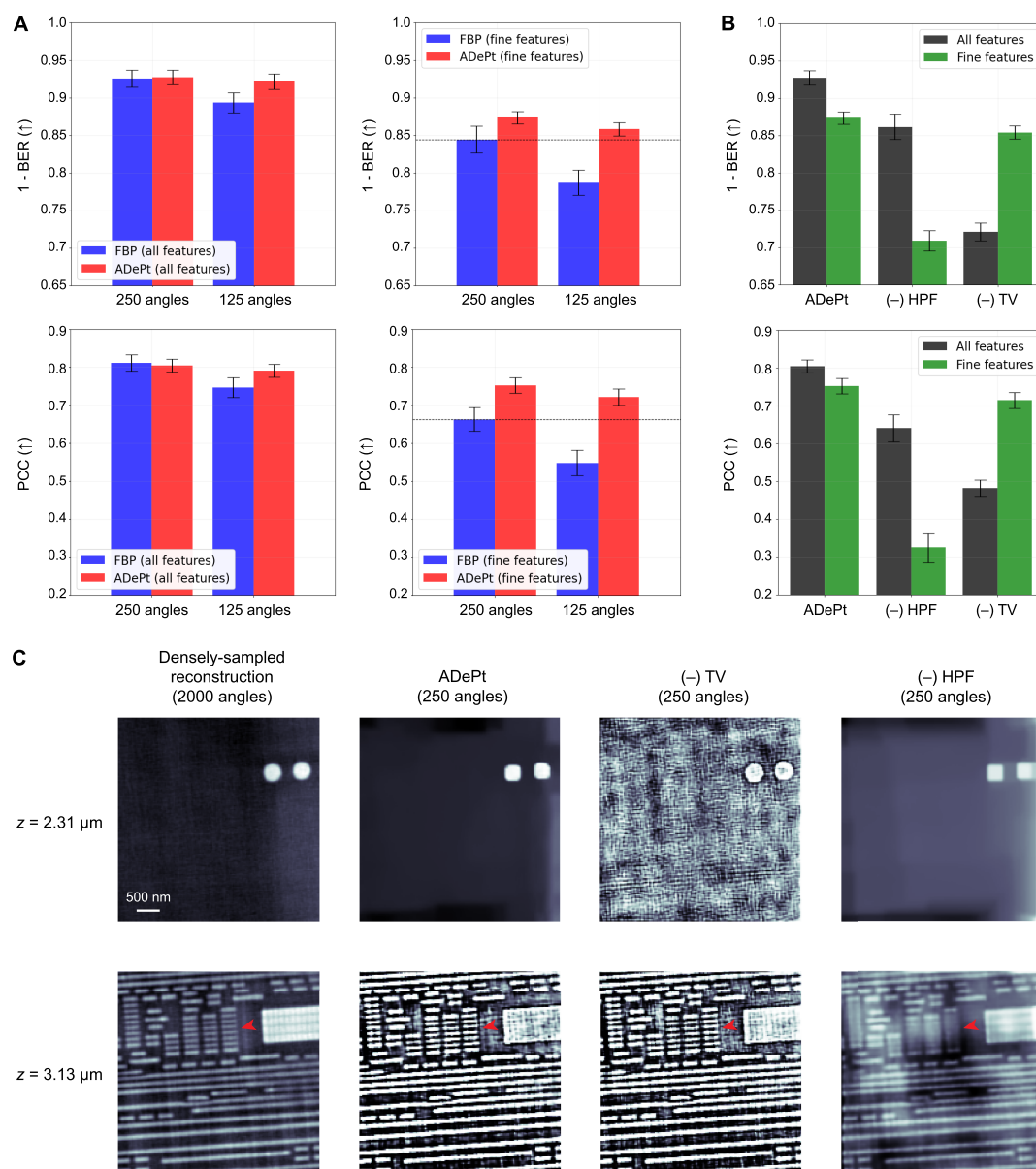
Fig. 5. Quantitative analysis of reconstructions and ablation study. (A) We use bit-error rate (BER) and Pearson correlation coefficient (PCC) to compare ADePt reconstructions with the baseline for different scales of features in the integrated circuits. (B) Ablation study. We assess relative contribution of each design element, i.e., high-pass filtering (HPF) and total-variation (TV) regularization, to the final reconstruction by removing one at a time from the complete model. We incorporate HPF to enforce a high-frequency content bias to our deep neural network and TV regularization to suppress spurious high-frequency artifacts in the background. Figure S4 in Supplement 1 illustrates (A) and (B) further using another quantitative metric. (C) The self-supervised learning algorithm behaves unfavorably when each component is ablated. TV regularization suppresses high-frequency artifacts, and HPF improves the spatial resolution of features recovered by the algorithm (red arrows).

REFERENCES

1. L. Helfen, T. Baumbach, P. Mikulik, D. Kiel, P. Pernot, P. Cloetens, and J. Baruchel, "High-resolution three-dimensional imaging of flat objects by synchrotron-radiation computed laminography," Appl. Phys. Lett. **86**, 071915 (2005).

2. L. Helfen, A. Myagotin, P. Mikulík, P. Pernot, A. Voropaev, M. Di Michiel, J. Baruchel, and T. Baumbach, "On the implementation of computed laminography using synchrotron radiation," Rev. Sci. Instrum. **82**, 063702 (2011).

3. Z. H. Levine, A. R. Kalukin, S. P. Frigo, I. McNulty, and M. Kuhn, "Tomographic reconstruction of an integrated circuit interconnect," Appl. Phys. Lett. **74**, 150–152 (1999).

4. A. Tkachuk, M. Feser, H. Cui, F. Duewer, H. Chang, and W. Yun, "High-resolution x-ray tomography using laboratory sources," Proc. SPIE **6318**, 431–438 (2006).

5. M. Holler, A. Diaz, M. Guizar-Sicairos, P. Karvinen, E. Färm, E. Härkönen, M. Ritala, A. Menzel, J. Raabe, and O. Bunk, "X-ray ptychographic computed tomography at 16 nm isotropic 3D resolution," Sci. Rep. **4**, 3857 (2014).

6. M. Holler, M. Guizar-Sicairos, E. H. Tsai, R. Dinapoli, E. Müller, O. Bunk, J. Raabe, and G. Aeppli, "High-resolution non-destructive three-dimensional imaging of integrated circuits," Nature **543**, 402–406 (2017).

7. W. Hoppe, "Beugung im inhomogenen primärstrahlwellenfeld. i. prinzip einer phasenmessung von elektronenbeugungsinterferenzen," Acta Crystallogr. A **25**, 495–501 (1969).

8. R. Hegerl and W. Hoppe, "Dynamische Theorie der Kristallstrukturanalyse durch Elektronenbeugung im inhomogenen Primärstrahlwellenfeld," Berichte Bunsenges physikalische Chem. **74**, 1148–1154 (1970).

9. J. Rodenburg and R. Bates, "The theory of super-resolution electron microscopy via Wigner-distribution deconvolution," Philos. Trans. R. Soc. London A **339**, 521–553 (1992).

10. M. Guizar-Sicairos and P. Thibault, "Ptychography: a solution to the phase problem," Phys. Today **74**(9), 42–48 (2021).

11. J. M. Rodenburg and H. M. Faulkner, "A phase retrieval algorithm for shifting illumination," Appl. Phys. Lett. **85**, 4795–4797 (2004).

12. J. Rodenburg, A. Hurst, and A. Cullis, "Transmission microscopy without lenses for objects of unlimited size," Ultramicroscopy **107**, 227–231 (2007).

13. H. M. L. Faulkner and J. Rodenburg, "Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm," Phys. Rev. Lett. **93**, 023903 (2004).

14. P. Thibault, M. Dierolf, O. Bunk, A. Menzel, and F. Pfeiffer, "Probe retrieval in ptychographic coherent diffractive imaging," Ultramicroscopy **109**, 338–343 (2009).

15. M. Odstrčil, A. Menzel, and M. Guizar-Sicairos, "Iterative least-squares solver for generalized maximum-likelihood ptychography," Opt. Express **26**, 3108–3123 (2018).

16. P. C. Konda, L. Loetgering, K. C. Zhou, S. Xu, A. R. Harvey, and R. Horstmeyer, "Fourier ptychography: current applications and future promises," Opt. Express **28**, 9603–9630 (2020).

17. G. Zheng, R. Horstmeyer, and C. Yang, "Wide-field, high-resolution Fourier ptychographic microscopy," Nat. Photonics **7**, 739–745 (2013).

18. L. Tian, X. Li, K. Ramchandran, and L. Waller, "Multiplexed coded illumination for Fourier ptychography with an LED array microscope," Biomed. Opt. Express **5**, 2376–2389 (2014).

19. M. Holler, M. Odstrcil, M. Guizar-Sicairos, M. Lebugle, E. Müller, S. Finizio, G. Tinti, C. David, J. Zusman, W. Unglaub, O. Bunk, J. Raabe, A. F. J. Levi, and G. Aeppli, "Three-dimensional imaging of integrated circuits with macro-to nanoscale zoom," Nat. Electron. **2**, 464–470 (2019).

20. M. Holler, M. Odstrčil, M. Guizar-Sicairos, M. Lebugle, U. Frommherz, T. Lachat, O. Bunk, J. Raabe, and G. Aeppli, "Lamni–an instrument for x-ray scanning microscopy in laminography geometry," J. Synchrotron. Radiat. **27**, 730–736 (2020).

21. I. Kang, Y. Jiang, M. Holler, M. Guizar-Sicairos, A. F. J. Levi, J. Klug, S. Vogt, and G. Barbastathis, "ADePt," GitHub (2023), https://github.com/iksungk/ADePt.

22. M. Guizar-Sicairos, I. Johnson, A. Diaz, M. Holler, P. Karvinen, H.-C. Stadler, R. Dinapoli, O. Bunk, and A. Menzel, "High-throughput ptychography using Eiger: scanning x-ray nano-imaging of extended regions," Opt. Express **22**, 14859–14870 (2014).

23. K. Wakonig, H.-C. Stadler, M. Odstrčil, E. H. Tsai, A. Diaz, M. Holler, I. Usov, J. Raabe, A. Menzel, and M. Guizar-Sicairos, "Ptychoshelves, a versatile high-level framework for high-performance analysis of ptychographic data," J. Appl. Crystallogr. **53**, 574–586 (2020).

24. M. Odstrčil, M. Holler, J. Raabe, and M. Guizar-Sicairos, "Alignment methods for nanotomography with deep subpixel accuracy," Opt. Express **27**, 36637–36652 (2019).

25. D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2018), pp. 9446–9454.

26. Y. Gandelsman, A. Shocher, and M. Irani, ""Double-DIP": unsupervised image decomposition via coupled deep-image-priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2019), pp. 11026–11035.

27. G. Mataev, P. Milanfar, and M. Elad, "Deepred: deep image prior powered by red," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (2019).

28. F. Wang, Y. Bian, H. Wang, M. Lyu, G. Pedrini, W. Osten, G. Barbastathis, and G. Situ, "Phase imaging with an untrained neural network," Light Sci. Appl. **9**, 1 (2020).

29. E. Bostan, R. Heckel, M. Chen, M. Kellman, and L. Waller, "Deep phase decoder: self-calibrating phase microscopy with an untrained deep neural network," Optica **7**, 559–562 (2020).

30. K. Gong, C. Catana, J. Qi, and Q. Li, "Pet image reconstruction using deep image prior," IEEE Trans. Med. Imaging **38**, 1655–1665 (2018).

31. D. O. Baguer, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," Inverse Prob. **36**, 094004 (2020).

32. J. Liu, Y. Sun, C. Eldeniz, W. Gan, H. An, and U. S. Kamilov, "Rare: image reconstruction using deep priors learned without groundtruth," IEEE J. Sel. Top. Signal Process. **14**, 1088–1099 (2020).

33. V. Sitzmann, J. Martel, A. Bergman, D. Lindell, and G. Wetzstein, "Implicit neural representations with periodic activation functions," in *Advances in Neural Information Processing Systems* (2020), Vol. **33**, pp. 7462–7473.

34. Y. Sun, J. Liu, M. Xie, B. Wohlberg, and U. S. Kamilov, "Coil: Coordinate-based internal learning for tomographic imaging," IEEE Trans. Comput. Imaging **7**, 1400–1412 (2021).

35. B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, and R. Ng, "Nerf: representing scenes as neural radiance fields for view synthesis," Commun. ACM **65**, 99–106 (2021).

36. R. Liu, Y. Sun, J. Zhu, L. Tian, and U. S. Kamilov, "Recovery of continuous 3D refractive index maps from discrete intensity-only measurements using neural fields," Nat. Mach. Intell. **4**, 781–791 (2022).

37. O. Ronneberger, P. Fischer, and T. Brox, "U-net: convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, 2015), pp. 234–241.

38. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the 13th International Conference on Artificial Intelligence and Statistics* (JMLR Workshop and Conference Proceedings, 2010), pp. 249–256.

39. S. Li and G. Barbastathis, "Spectral pre-modulation of training examples enhances the spatial resolution of the phase extraction neural network (PHENN)," Opt. Express **26**, 29340–29352 (2018).

40. I. Daubechies, *Ten Lectures on Wavelets* (SIAM, 1992).

41. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," arXiv, arXiv:1412.6980 (2014).

42. K. C. Zhou and R. Horstmeyer, "Diffraction tomography with a deep image prior," Opt. Express **28**, 12872–12896 (2020).

43. I. Kang, Z. Wu, Y. Jiang, Y. Yao, J. Deng, J. Klug, S. Vogt, and G. Barbastathis, "Attentional ptycho-tomography (APT) for three-dimensional nanoscale x-ray imaging with minimal data acquisition and computation time," Light Sci. Appl. **12**, 131 (2023).

44. I. Kang, A. Goy, and G. Barbastathis, "Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views," Light Sci. Appl. **10**, 74 (2021).

45. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," arXiv, arXiv:1706.03762 (2017).

46. H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, "Axial-deeplab: stand-alone axial-attention for panoptic segmentation," in *European Conference on Computer Vision* (Springer, 2020), pp. 108–126.

47. L. Tian, J. C. Petruccelli, and G. Barbastathis, "Nonlinear diffusion regularization for transport of intensity phase imaging," Opt. Lett. **37**, 4131–4133 (2012).