

ARTICLE

Open Access

Attentional Ptycho-Tomography (APT) for three-dimensional nanoscale X-ray imaging with minimal data acquisition and computation time

Iksung Kang^{1,5}, Ziling Wu^{2,6}, Yi Jiang³, Yudong Yao^{3,7}, Junjing Deng³, Jeffrey Klug³, Stefan Vogt³ and George Barbastathis^{2,4}✉

Abstract

Noninvasive X-ray imaging of nanoscale three-dimensional objects, such as integrated circuits (ICs), generally requires two types of scanning: ptychographic, which is translational and returns estimates of the complex electromagnetic field through the IC; combined with a tomographic scan, which collects these complex field projections from multiple angles. Here, we present Attentional Ptycho-Tomography (APT), an approach to drastically reduce the amount of angular scanning, and thus the total acquisition time. APT is machine learning-based, utilizing axial self-Attention for Ptycho-Tomographic reconstruction. APT is trained to obtain accurate reconstructions of the ICs, despite the incompleteness of the measurements. The training process includes regularizing priors in the form of typical patterns found in IC interiors, and the physics of X-ray propagation through the IC. We show that APT with $\times 12$ reduced angles achieves fidelity comparable to the gold standard Simultaneous Algebraic Reconstruction Technique (SART) with the original set of angles. When using the same set of reduced angles, then APT also outperforms Filtered Back Projection (FBP), Simultaneous Iterative Reconstruction Technique (SIRT) and SART. The time needed to compute the reconstruction is also reduced, because the trained neural network is a forward operation, unlike the iterative nature of these alternatives. Our experiments show that, without loss in quality, for a $4.48 \times 93.2 \times 3.92 \mu\text{m}^3$ IC ($\approx 6 \times 10^8$ voxels), APT reduces the total data acquisition and computation time from 67.96 h to 38 min. We expect our physics-assisted and attention-utilizing machine learning framework to be applicable to other branches of nanoscale imaging, including materials science and biological imaging.

Introduction

Three-dimensional X-ray imaging enables noninvasive monitoring of objects' interiors with nanoscale resolution. Integrated circuits (IC) are especially interesting for this operation, for two reasons: first, noninvasive inspection of ICs is important for verifying manufacturing integrity. Second, ICs follow specific design rules, which makes their geometries highly regular and yet highly diverse. The

geometrical properties are then useful as prior knowledge, enabling vast improvements in practical aspects of the imaging process, such as acquisition time as we show here.

Prior works have typically used two types of scanning: translational and rotational. The translational scan (*ptycho*) is inspired by ptychography, i.e. a scanning-based coherent diffraction imaging method for phase retrieval. Ptychography was originally proposed by W. Hoppe¹ to solve the phase problem in Scanning Transmission Electron Microscopy (STEM), where a moving aperture resolves the ambiguity in phase based on translational invariance. The term “ptychography” was coined in the following year². Nellist et al.³ demonstrated resolution

Correspondence: George Barbastathis (gbarb@mit.edu)

¹Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Full list of author information is available at the end of the article

These authors contributed equally: Iksung Kang, Ziling Wu

© The Author(s) 2023



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.



improvement in STEM by a factor of 2.5 over the limit imposed by partial coherence, exploiting the redundancy in the ptychographic measurements. As an alternative that does not even require careful aberration correction in the optics, Gerchberg and Saxton⁴ introduced a lensless iterative phase retrieval algorithm, now referred to as *GS* after them. This work was extended to lensless ptychography for extended objects by Faulkner⁵. Subsequently, Rodenburg⁶ introduced yet another iterative phase retrieval algorithm called Ptychographical Iterative Engine (PIE) that simultaneously retrieves both the object and the probe function. Thus, the requirement of a high-quality lens for imaging is fundamentally lifted. Further advances by Thibault et al.⁷ and Thibault and Guizar-Sicairos⁸ led to the Difference Map (DM) algorithm and Maximum Likelihood algorithm, respectively, for iterative ptychographic reconstruction.

After the ptychographic reconstruction step, the second angular scan (*tomo*) operation is required to retrieve the object's interior, as in tomography. For parallel beam illumination and under the weak scattering approximation, the measurements are interpreted simply as projections through the object, i.e. the measurements implement the object's Radon transform^{9,10}. The inverse Radon transform is typically implemented as a version of the Filtered BackProjection (FBP) algorithm, first proposed by Bracewell and Riddle^{11,12}. Gordon, Bender, and Herman¹³ proposed an alternative iterative tomography algorithm called Algebraic Reconstruction Technique (ART), which applies also to non-parallel illumination beams and works by updating the object estimate to sequentially bring each reconstructed projection into agreement with the corresponding measured projection. Subsequent improvements of this original iterative method were the Simultaneous Iterative Reconstruction Technique (SIRT)¹⁴ and the Simultaneous Algebraic Reconstruction Technique (SART)¹⁵, which consider all projections simultaneously and thus drastically reduce the number of iterations for the reconstruction process. Maximum Likelihood methods have also been popular for tomography, with the Bouman-Sauer algorithm¹⁶ as one of the most prominent.

For X-rays, the high penetration depth facilitates recovery of information deep inside the sample in the angular sampling scheme. Combining this property with translational scanning for lensless high spatial resolution, Dierolf et al.¹⁷ proposed the Ptychographic X-ray Computed Tomography (PXCT) scheme to determine the volumetric interior of biological specimens with nanoscale details. Using this technique, Holler et al.¹⁸ experimentally demonstrated noninvasive imaging of ICs produced with 22-nm technology at 14.6-nm resolution. PXCT provides noninvasive inspection of fabricated samples, eliminating the need for destructive measures such as delayering,

which are required by STEM due to its limited depth access caused by electron scattering. This allows fabs to connect to synchrotron X-ray sources and achieve quality assurance without destructive measures. However, this technique is limited by the requirement for two types of scanning, angular and translational, and scale badly scales with object volume. A novel X-ray microscope called Velociprobe¹⁹ utilizes fly-scan ptychography²⁰ to significantly reduce the data acquisition time. Still, total data acquisition and reconstruction time for a typical $100 \times 100 \times 5 \mu\text{m}^3$ IC ($\approx 2 \times 10^{10}$ voxels) is estimated to be in excess of two months.

Here, we propose a machine learning framework to reduce data acquisition and computation time for IC reconstruction under the X-ray ptycho-tomography geometry to provide a noninvasive and efficient solution for inspection purposes. The reduction in data acquisition is compensated by explicit use of prior knowledge of the typical objects being imaged, and of the optical physics of the imaging system. The length of the acquisition and computation time scale as the number N of tomo-scans. The total angular range θ determines the size of the missing wedge in the Fourier domain and, therefore, is commensurate with loss of fidelity. Our “gold standard” is a ptycho-tomo reconstruction by SART with $N = 349$ and $\theta = \pm 70.4^\circ$. This maximum angle is determined by practical considerations, such as the sample geometry. More details about the gold standard geometry and our approach are available in “Materials and methods” section.

To search this two-dimensional space (N, θ), our strategy is as follows: we start with the gold standard nominal values of N and θ . If we reduced N while using a standard reconstruction algorithm, like FBP, SIRT, etc. mentioned earlier, performance would decrease immediately. With machine learning, we find that it is possible to regularize for the loss of angular sampling density and still maintain reconstruction fidelity, down to a minimum number N^* . Then we start reducing the total angular range, meaning that the sampling now becomes denser. The machine-learning regularizer again manages to maintain approximately even fidelity down to a minimum range θ^* . This is our optimal operating point (N^*, θ^*). The strategy is depicted in Fig. 1b. In principle, this procedure can be repeated to find even tighter operating conditions, but we did not carry that out as we would expect any further gains to be minimal.

That machine learning is suitable for achieving even fidelity while the amount of sampled data is decreasing is not entirely surprising by now. The key is the ability of deep neural networks to very effectively capture regularizing priors, especially sparsity, in both supervised mode as we do here and in untrained mode^{21,22}. Previous demonstrations of supervised learning have been carried



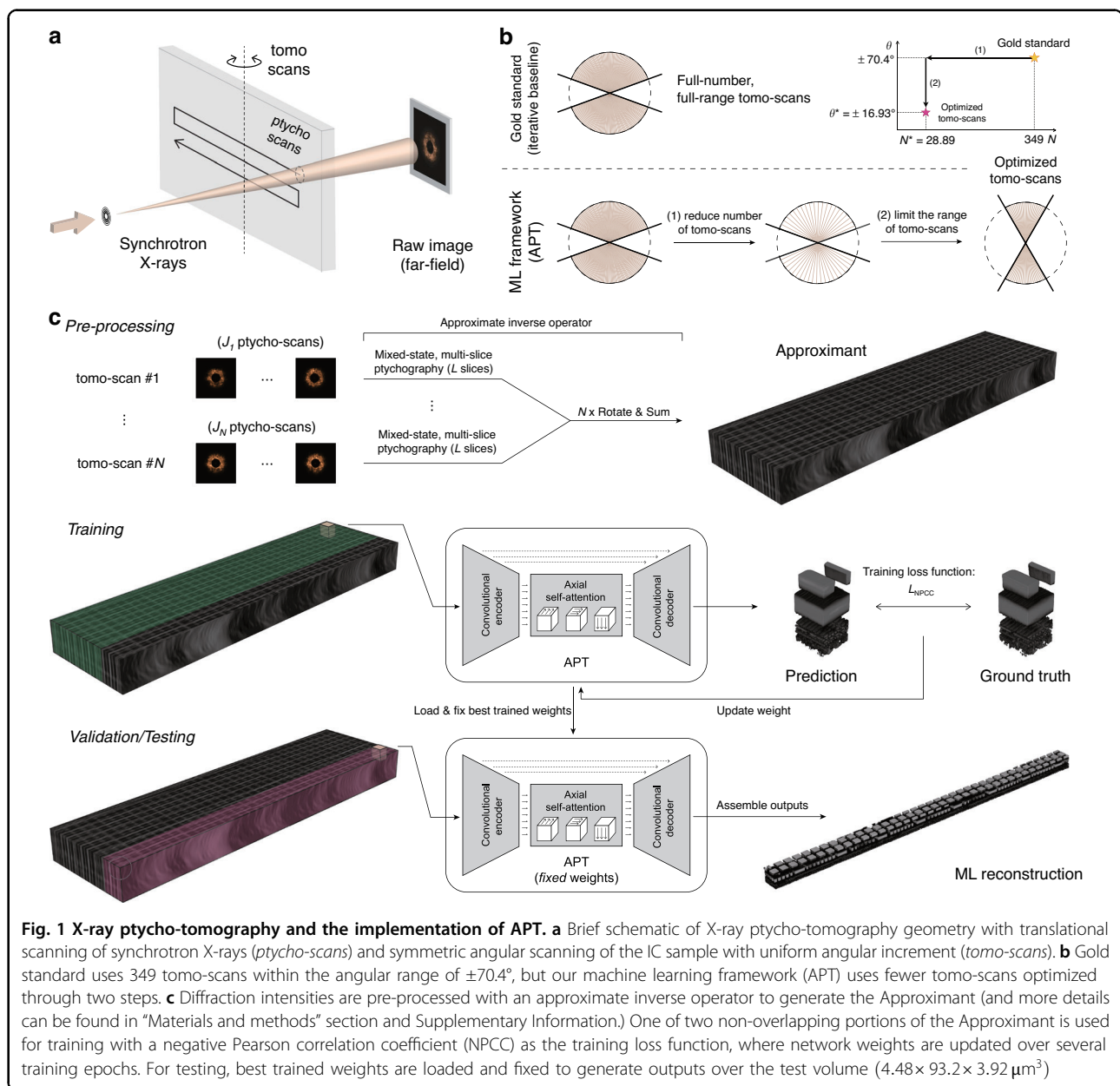


Fig. 1 X-ray ptycho-tomography and the implementation of APT. **a** Brief schematic of X-ray ptycho-tomography geometry with translational scanning of synchrotron X-rays (ptycho-scans) and symmetric angular scanning of the IC sample with uniform angular increment (tomo-scans). **b** Gold standard uses 349 tomo-scans within the angular range of $\pm 70.4^\circ$, but our machine learning framework (APT) uses fewer tomo-scans optimized through two steps. **c** Diffraction intensities are pre-processed with an approximate inverse operator to generate the Approximant (and more details can be found in “Materials and methods” section and Supplementary Information.) One of two non-overlapping portions of the Approximant is used for training with a negative Pearson correlation coefficient (NPCC) as the training loss function, where network weights are updated over several training epochs. For testing, best trained weights are loaded and fixed to generate outputs over the test volume ($4.48 \times 93.2 \times 3.92 \mu\text{m}^3$)

out for Fourier ptychography^{23,24} and two-dimensional ptychography²⁵. The reason we chose the supervised learning mode is because we had ample data available from the gold standard ptycho-tomo approach.

APT is described schematically in Fig. 1c. We first invert the far-field diffraction intensities (or ptycho-scans) with an approximate inversion operator. This yields to get an approximate volumetric estimate of the interior of an IC chip, which we dub the “Approximant”²⁶. This step utilizes prior knowledge on underlying imaging physics and pre-processes the input with the physics prior. The Approximant as a result of this pre-processing step (or physics-informing step) is defective in terms that layers

are not well separated because of the approximate inversion from diffraction intensities and only a small fraction of tomo-scans used for the computation. During training, the neural network’s weights are optimized based on the Approximant as input. Upon the completion, the trained neural network gives a refined volumetric reconstruction of ICs.

The proposed neural network is based on a 3D U-net architecture^{27,28} and augmented with the multi-head axial self-attention²⁹ to address lack of spatial resolution in the Approximant by taking advantage of its global-range interactions to retrieve information from all layers to resolve each layer’s structure. We choose this multi-head



axial self-attention over multi-head self-attention³⁰ to alleviate computational burden.

We demonstrate that the present method is capable of providing reliable reconstructions of ICs even when both the number and the total angular range of tomo-scans are largely decreased to $N^* \sim 29$ and $\theta^* \sim \pm 17^\circ$ representing an improvement of $\times 12$ and $\times 4.2$, respectively. For the reconstruction of an IC chip over the test volume ($4.48 \times 93.2 \times 3.92 \mu\text{m}^3$), 0.63 h (or 38 min) is sufficient for both data acquisition and reconstruction with our machine learning framework. The improvements work out to an approximate overall $\times 108$ reduction in total (acquisition plus computation) time compared to the current up-to-date iterative reconstruction method. We are confident that this method can be applied to various physical systems in which the forward model can be expressed mathematically. Furthermore, this method is not limited to specific sample geometries and can be extended to other types of samples.

Results

Reducing acquisition and scanning time

The synchrotron beam is delivered on the sample, and a full lateral scan is carried out to obtain the ptychographic information for each angular orientation of the sample. Repeating for N angles collects tomographic information for the interior's reconstruction. The raw intensities past the sample are recorded by a digital camera detector at each scan position. The details of the experimental collection system are in "Materials and methods" section. The collected raw intensities are then processed in two steps: the first step embeds the physics of X-ray propagation through an Approximant operator^{31,32}, while the second step consists of the APT network delivering the final reconstruction, as described earlier. The details of training and operating this computational pipeline are in "Materials and methods" section. As discussed earlier, our approach is to first reduce scanning time by finding the minimum N^* and then reduce computation time by finding the minimum θ^* . While we acknowledge that optimizing both parameters jointly for a few iterations would be ideal, we believe that the resulting approximant is only a preliminary input to our neural network. Therefore, it is unnecessary to overly complicate its derivation, and keeping it simple would be more beneficial.

A parameter sweep over N is shown qualitatively in Fig. 2. Four quantitative performance comparisons are shown in Fig. 3, in terms of the following metrics: Pearson correlation coefficient (PCC)³³, multi-scale structural similarity index metric (MS-SSIM)³⁴, Dice Similarity coefficient (DSC)³⁵, and Bit-Error Rate (BER, more details available in "Materials and methods"

section). Both analyses indicate that $N^* \sim 29$, representing a reduction of more than $\times 12$ over the gold standard of $N = 349$. Reducing N significantly below this value results in noticeable degradation, both qualitatively and quantitatively.

Next we fix $N = 29$ and perform a parameter sweep over θ . Qualitative results are shown in Fig. 4, while the quantitative evaluation according to the same four metrics of the previous section is in Fig. 5. Both analyses lead to $\theta^* \sim \pm 17^\circ$ as the approximate lower bound before drastic degradation occurs. While the number of tomo-scans (N) is the primary determinant of computation time, the angular range of tomo-scans θ can also have an impact. This is because the depth of multi-slice estimates $O_r^{(n)[l]}$ derived from different tomographic angles varies with the angle, as the optical path length changes accordingly. This can affect the speed of computing the Approximant. More information is provided in the "Materials and methods" section. The savings in data acquisition and computation times are $\times 12$ and $\times 105$, respectively, and total time savings (acquisition plus computation) of $\times 108$.

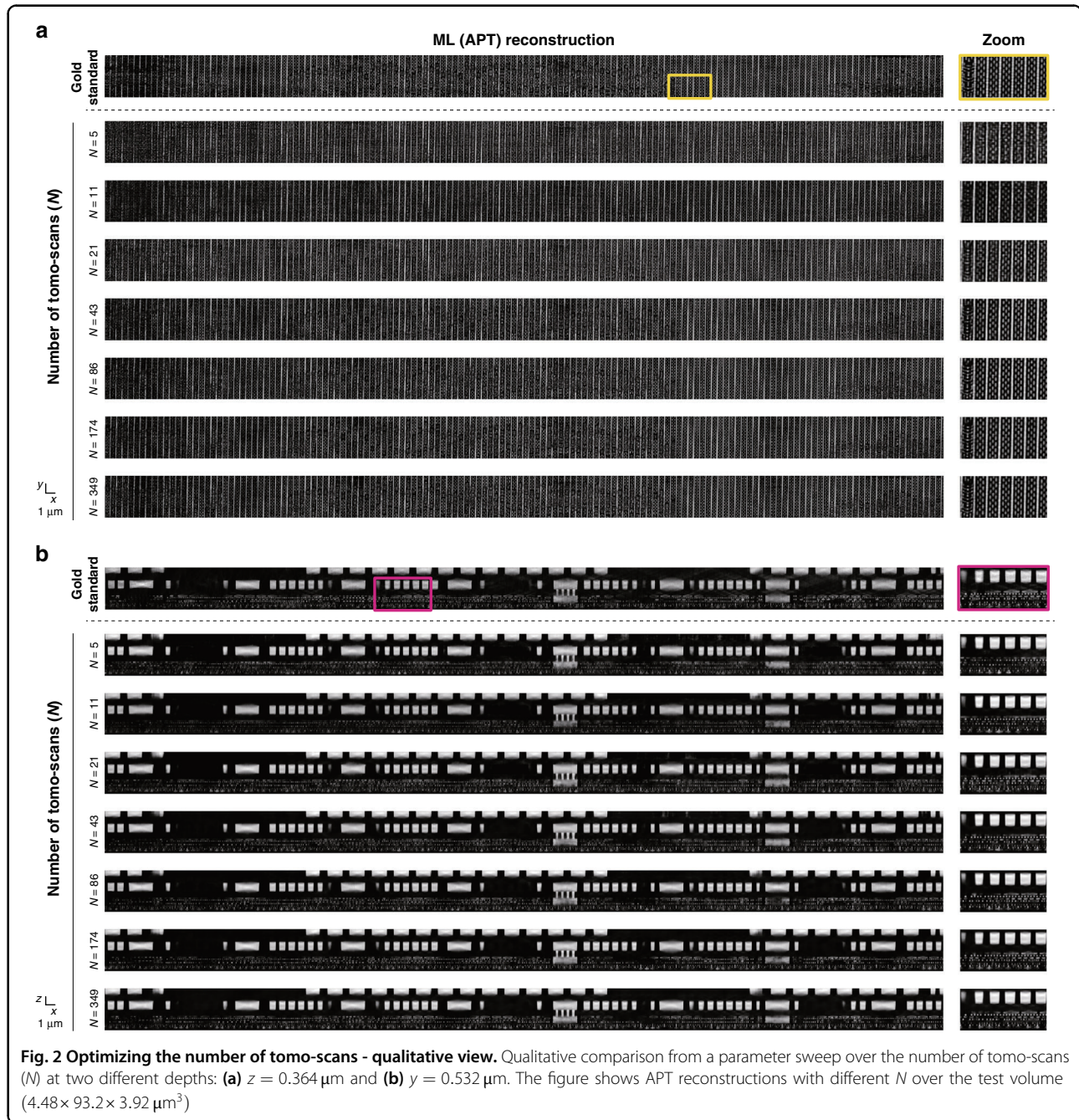
Regularization and imaging system physics

The reported improvements suggest that the APT algorithm is particularly effective at learning regularizing priors to compensate for the missing information. Figure 6a shows the power spectral densities of the gold standard, the APT reconstruction, and the baseline tomographic reconstruction methods FBP¹¹, SIRT¹⁴, and SART¹⁵, all obtained at $N^* = 29$ and $\theta^* = \pm 16.8^\circ$. The missing wedge is evident in the latter three. The qualitative cross-sections in Fig. 6b confirm that the missing wedge effect leads to severe artifacts in the baseline methods, but not in APT. Further discussions, including a comparison with TV-FISTA^{36,37}, are provided in the Supplementary Information.

APT also relies on its input, the Approximant, having carefully taken into account the physics of the imaging system. Unlike earlier works where the illumination on the sample was coherent^{31,32}, the synchrotron may be considered as temporally coherent but is rather less coherent spatially. The mutual intensity is expressed as a linear combination of mutually incoherent states, also known as coherent modes³⁸. Accounting correctly for the synchrotron X-ray's coherence state has been shown to improve spatial resolution and phase contrast in standard ptychography for thin samples³⁹.

For samples thicker than the depth of focus of the probe, multi-slice reconstruction from simple ptychography has been demonstrated with visible light⁴⁰, X-rays⁴¹, and electrons⁴². This is the starting point for our Approximant as shown Fig. 1c, and please find more information on why multi-slice ptychography was used instead of 2D ptychography in Gradient calculation in





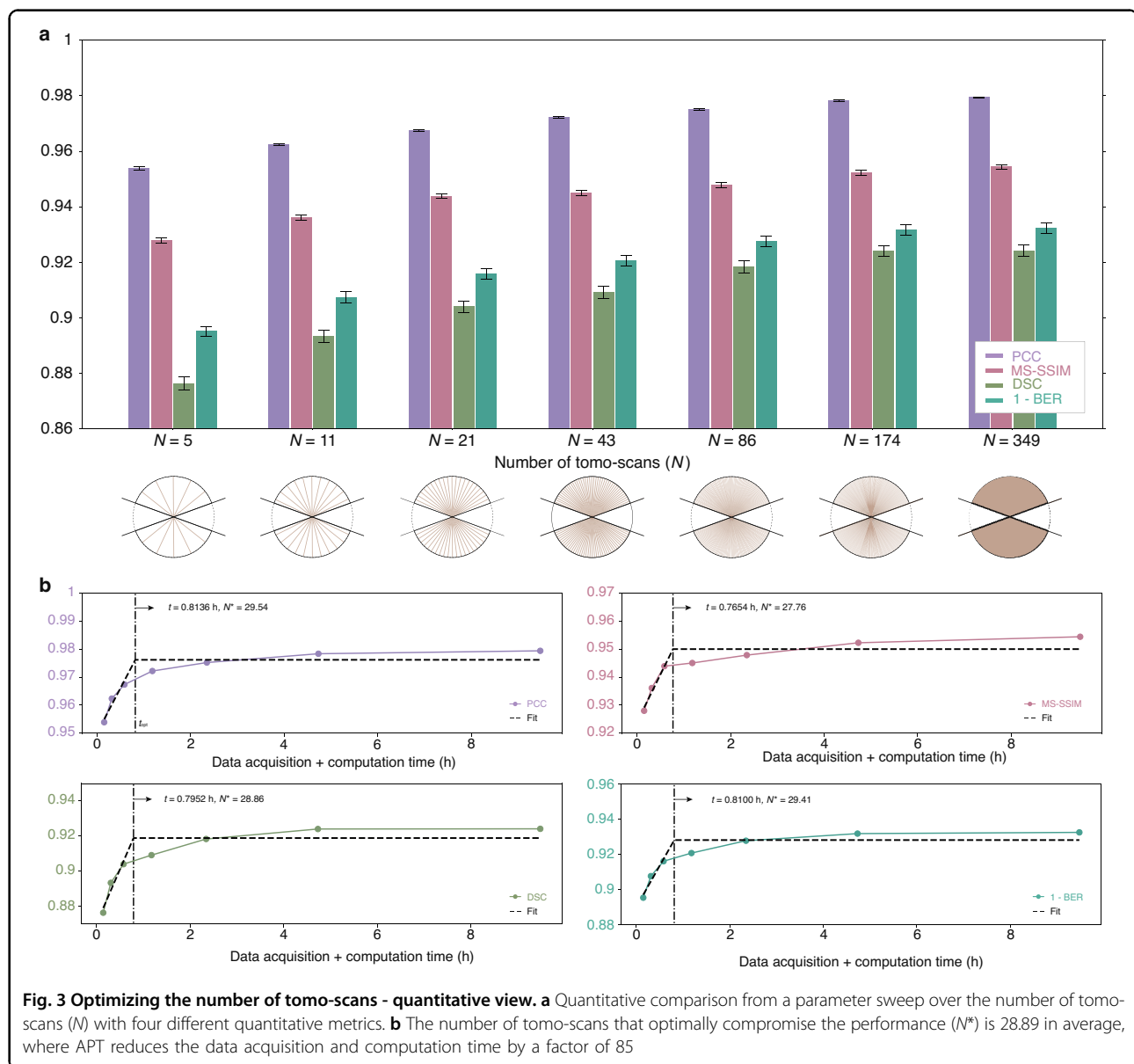
“Materials and methods” section. We form the cost function

$$\mathcal{L}_n = \sum_{j=1}^{J_n} \sum_q \left(\sqrt{\sum_{m=1}^M |P_d \{P_{j,r,m}^{(n)[L]} O_r^{(n)[L]}\}|^2} - \sqrt{I_{j,q}^{(n)}} \right)^2 \quad (n = 1, 2, \dots, N) \quad (1)$$

where N is the number of given tomo-scans; J_n the number of pycho-scans associated with the n -th tomo-scan; M the number of coherent modes; L the number of slices for our given depth of focus works out to equal to 5; q denotes the coordinates in the reciprocal space;

P_d the free-space propagator by the distance d ; and $P_{j,r,m}^{(n)[L]}, I_{j,q}^{(n)}$ indicate the wavefield before the L -th slice from the m -th coherent mode and the experimental diffraction intensity at the j -th pycho and the n -th tomo-scans, respectively. We run two iterations of a gradient scheme on Eq. 1 and obtain the argument $\angle O_r^{(n)[L]}$ at each one among the $l = 1, \dots, L$ slices^{40,42}. We rotate the result back to the original coordinate system, and average the estimates from all tomo-scan steps to yield the final Approximant. More details can be found in “Materials and methods” section.





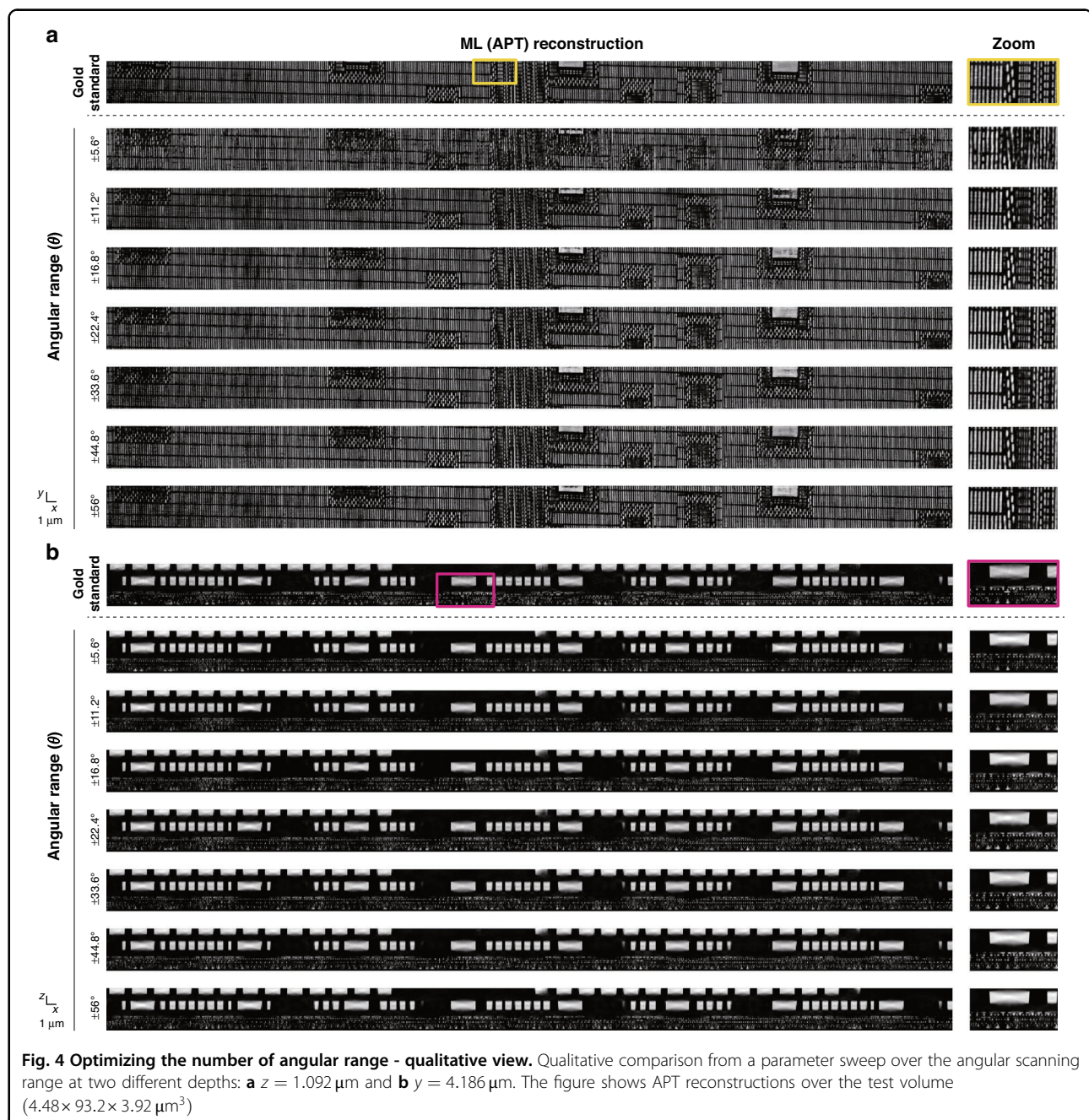
The Approximant computation step is the slowest in the pipeline; in our computing hardware (see Materials and Methods), it takes 36 min when $\theta = \pm 70.4^\circ$ and 26 min when $\theta \sim \pm 17^\circ$. In addition to the computation time, the spacing between slices in the Approximant is nominally limited by the depth of focus, and that is why we only reconstruct $L = 5$ of them. The number of desired reconstruction slices is much larger, i.e. 280, so we simply dilate the Approximant slices to match it. As a result, the input to the neural network is poor (more in Supplementary Information). Nevertheless, the subsequent APT architecture learns how to use the multi-slices as input and, as long as $N > N^*$ and $\theta > \theta^*$, produce a high-fidelity final reconstruction with much finer slice spacing.

Discussion

APT is trained using the gold standard reconstructions of randomly selected segments from a single IC specimen, which was made available for our experiments. This prompts us to address two related concerns: (1) what can we guarantee about fidelity of the gold standard and, hence, our reconstructions vis-à-vis the ground truth, i.e. the physical specimen? and (2) is our APT overtrained to this specific IC?

The first concern was partially addressed by refs. ^{31,32}, where the design files of the geometrical features were treated as ground truth. (That method was still bound by the assumption that the physical specimen matched the design files; but that was less of a concern, given the size scales involved.) Neither of these algorithms would have

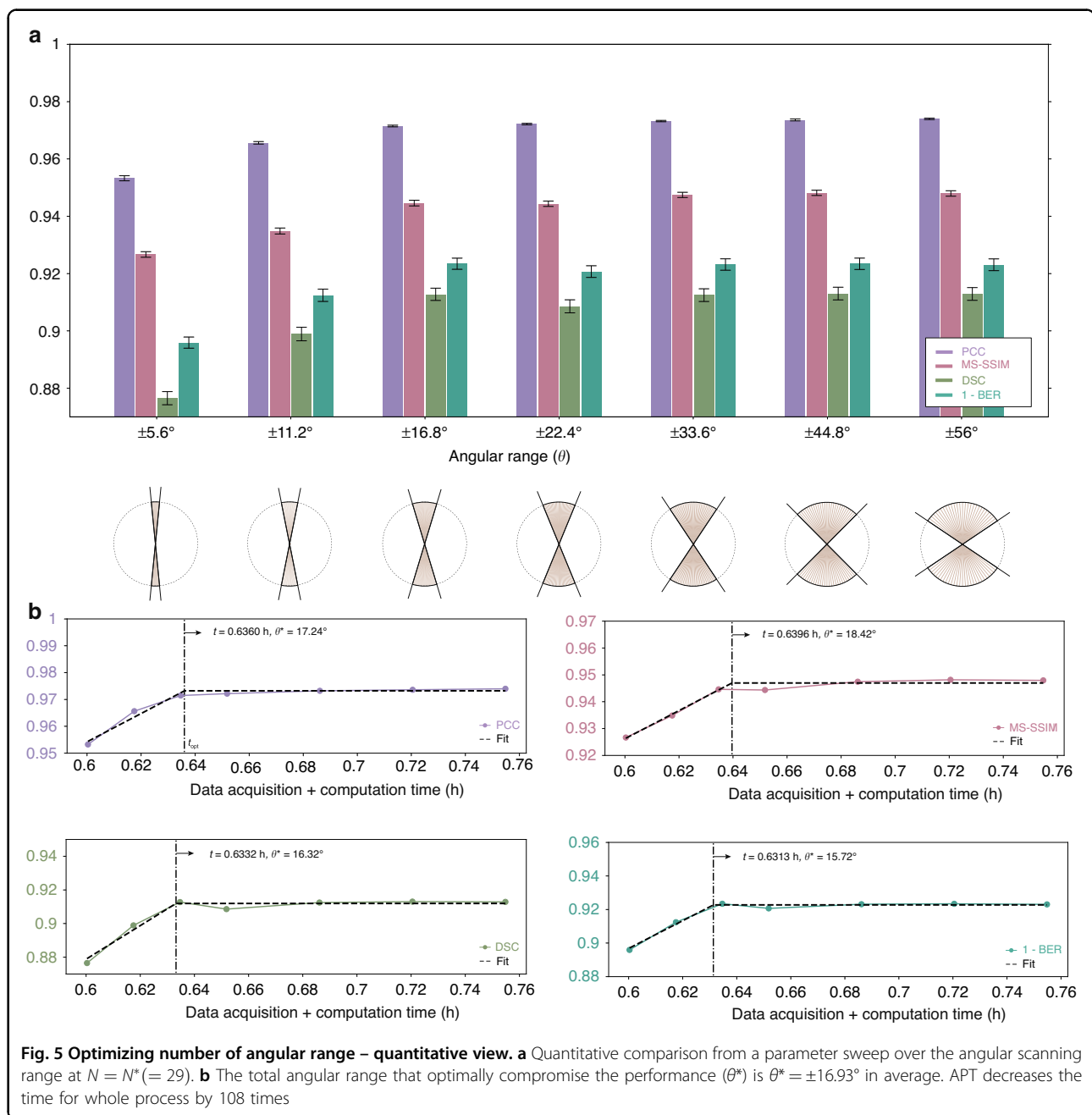




worked in the case reported here, because of the great range of feature sizes in the specimen and because the synchrotron X-ray is not spatially coherent. Moreover, the design files for the specimen are not available to us. On the other hand, the gold standard was obtained quite thoroughly with the scan step size of 100 nm in the pycho-scan and $N = 349$ angles in the tomo-scan. Besides, there are no discernible visual artifacts in the gold standard reconstructions. These facts provide us with reasonable assurance about the fairness of our comparisons in Figs. 2–6.

Regarding the second concern: if new structures are given where the priors are *significantly* different than the priors learnt here (e.g. features oriented at 45°) then APT would have to be retrained. This is a necessary limitation of our supervised learning approach. The same holds for non-IC objects such as viruses. If, moreover, not enough physical specimens are available for supervised training, then it is possible to train by rigorously simulating the forward propagation of X-rays through the specimen (as refs. ^{31,32} did for visible light) or use “untrained” methods, such as deep image prior^{21,43,44}.





The reported best values of $N^* \sim 29$ and $\theta^* \sim \pm 17^\circ$ are not fundamental, but indicative of the effectiveness of IC geometries acting as regularizing prior. Less complex geometries, smooth and with less content at high frequencies in the missing wedge, could achieve even better reductions, whereas complex structures with smaller features and higher refractive index contrast would be more limited. A full theoretical analysis of how N^* and θ^* depend on the complexity of the prior is beyond the scope of this work.

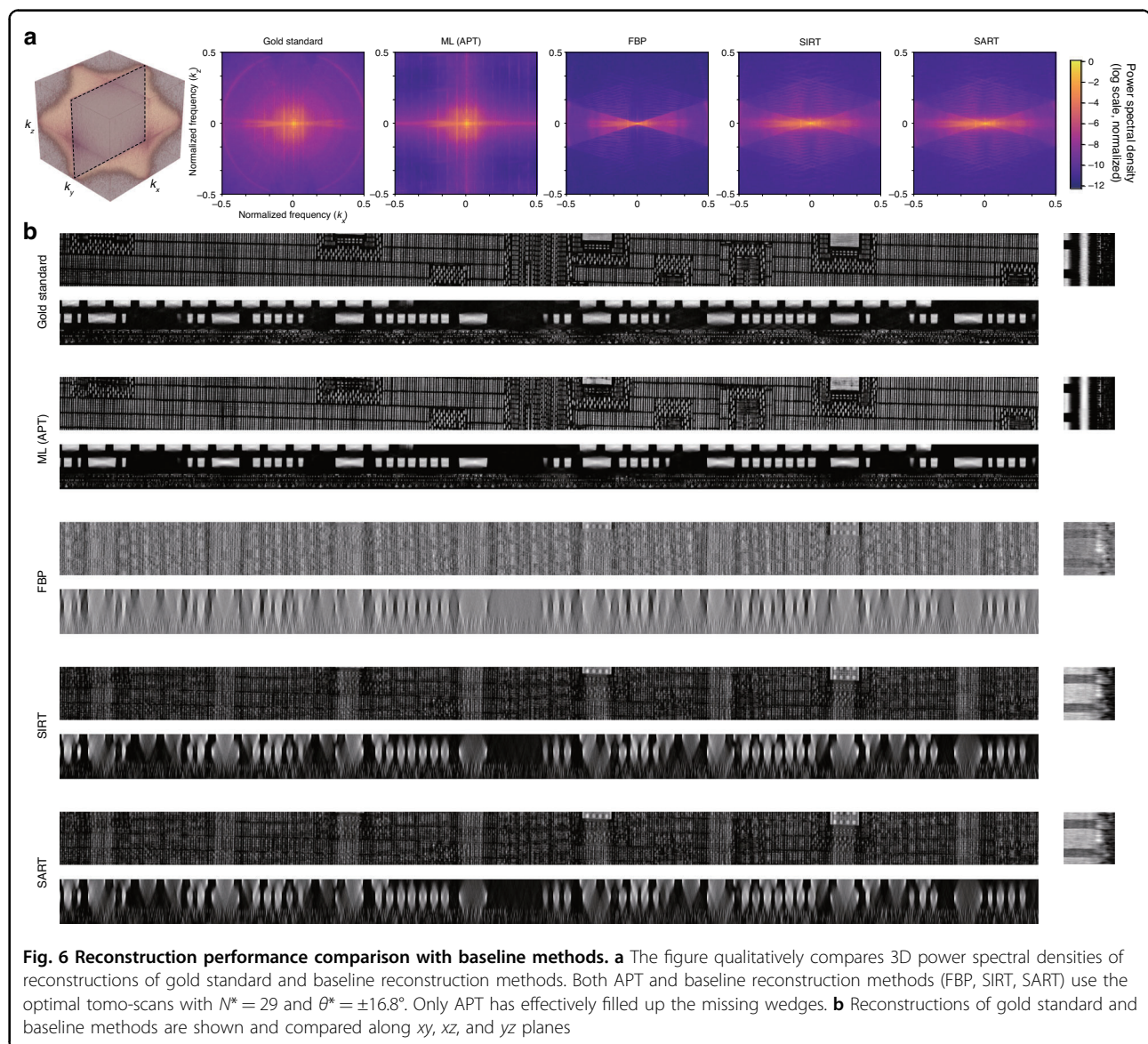
Lastly, regarding ICs in particular and planar samples more generally, the total attenuation of the X-rays increases at large

angles, which leads to artifacts. It may be compensated computationally, or by scanning the illumination wavevectors on a conical surface. The latter scheme is referred to as laminography^{45,46}. It is beyond the scope of our present work, but it would be interesting to investigate if approaches similar to the one reported here are applicable.

Materials and Methods

Experiment and the gold standard preparation

Integrated circuits produced with 16-nm technology of size $25.1 \times 93.2 \times 3.92 \mu\text{m}^3$ were laterally scanned with



synchrotron X-rays of 8.8 keV for each tomo scan with Velociprobe¹⁹ at the Advanced Photon Source (APS) of the Argonne National Laboratory (ANL). 12 coherent modes of the synchrotron X-ray were used for the experiment. Tomo-scans were carried out from -70.4° to 70.4° with angular increment of 0.4° , and for each tomo-scan, ptychoscans were recorded on-the-fly at ~ 60 k lateral positions with Dectris Eiger X 500 K area detector (pixel size: $75 \mu\text{m}$, sample-to-detector distance: 1.92 m) at a frame rate of 500 Hz with the scan step size of 100 nm . Elapsed time of this whole data acquisition process (translational and angular) was 12.51 h , or 129 s per tomo-scan.

As a first step to obtain the gold standard reconstruction, a two-dimensional projection was reconstructed for each tomo-scan with 600 iterations of the least-square

maximum likelihood ptychographic algorithm⁴⁷ as implemented in PtychoShelves⁴⁸. The ptychographic reconstruction for all 349 tomo-scans was processed with 8 Tesla V100 GPUs in parallel to expedite the process, thus taking 362.09 h for this step.

Then, the projections were aligned to a tomographic rotation axis with an additional correction in the form of a phase ramp removal process. Then, a deep neural network pre-trained on similar images of integrated circuits was applied to the aligned projections for upsampling by $\times 2$ ^{49,50}. The elapsed time of this step was approximately 5 h .

Lastly, the final tomographic reconstruction was performed using 349 upsampled projections with 10 iterations of SART to generate a finally three-dimensional reconstruction of the IC sample with the isotropic 14-nm voxel size, which took 1 hr with 8 Tesla V100 GPUs.



Gradient calculation

Considering the mixed-state (spatially partially coherent) nature of synchrotron X-rays and multi-slice structure of the IC sample, a forward model can be formulated as

$$\psi_{j,r,m}^{(n)[L]} = O_r^{(n)[L]} \mathcal{P}_{\Delta z} \left[O_r^{(n)[L-1]} \mathcal{P}_{\Delta z} \left[\cdots \mathcal{P}_{\Delta z} \left[P_{r-r_j,m} O_r^{(n)[1]} \right] \right] \right] \quad (2)$$

$$I_{j,q}^{(n)} = \sum_{m=1}^M \left| \tilde{\psi}_{j,q,m}^{(n)[L]} \right|^2 = \sum_{m=1}^M \left| \mathcal{F} \left[\psi_{j,r,m}^{(n)[L]} \right] \right|^2 \quad (3)$$

where

n : the index of tomo-scans ($n = 1, 2, \dots, N$)

j : the index of ptycho-scans ($j = 1, 2, \dots, J_n$)

l : the index of multi-slices ($l = 1, 2, \dots, L$)

m : the index of mixed states ($m = 1, 2, \dots, M$)

r : the spatial domain coordinates

q : the reciprocal domain coordinates

$P_{r,m}$: the m -th coherent mode of the synchrotron X-ray probe

$O_r^{(n)[l]}$: the l -th slice of the object viewed at the n -th tomo-scan.

The following describes the gradient computation of the loss function in Eq. 1 based on the forward model, which was done automatically with Ptychoshelves⁴⁸. The gradients of the loss function with respect to the wavefield and the complex object are

$$\frac{\partial \mathcal{L}}{\partial P_{j,r,m}^{(n)[l]}} = \begin{cases} O_r^{(n)[l]} \left(\mathcal{P}_{-\Delta z} \left(\frac{\partial \mathcal{L}}{\partial P_{j,r,m}^{(n)[l+1]}} \right) \right)^* & \text{for } 1 \leq l < L \\ O_r^{(n)[L]} \chi_{j,r,m}^{(n)[L]*} & \text{for } l = L \end{cases} \quad (4a)$$

and

$$\frac{\partial \mathcal{L}}{\partial O_r^{(n)[l]}} = \begin{cases} \sum_{j=1}^{J_n} \sum_{m=1}^M P_{j,r,m}^{(n)[l]} \left(\mathcal{P}_{-\Delta z} \left(\frac{\partial \mathcal{L}}{\partial P_{j,r,m}^{(n)[l+1]}} \right) \right)^* & \text{for } 1 \leq l < L \\ \sum_{j=1}^{J_n} \sum_{m=1}^M P_{j,r,m}^{(n)[L]} \chi_{j,r,m}^{(n)[L]*} & \text{for } l = L \end{cases} \quad (4b)$$

where

$$P_{j,r,m}^{(n)[l]} = \mathcal{P}_{\Delta z} \left[O_r^{(n)[l-1]} P_{j,r,m}^{(n)[l-1]} \right] \quad (6)$$

$$\frac{\partial \mathcal{L}}{\partial P_{r,m}} = \sum_{n=1}^N \sum_{j=1}^{J_n} \frac{\partial \mathcal{L}}{\partial P_{j,r+r_j,m}^{(n)[1]}} \quad (7)$$

$$\chi_{j,r,m}^{(n)[L]} = \mathcal{F}^{-1} \left\{ \left(1 - \frac{\sqrt{I_{j,q}^{(n)}}}{\left| \tilde{\psi}_{j,q,m}^{(n)[L]} \right|} \right) \tilde{\psi}_{j,q,m}^{(n)[L]} \right\} \quad (8)$$

With two iterations of gradient descent on the loss function in Eq. 1, we obtain the multi-slice estimate $O_r^{(n)[l]}|_{N_{\text{iter}}=2}$ for each tomo-scan and subsequently its argument at each one of the $L = 5$ slices. For the final

Approximant, we rotate the results back to the original coordinate system, and average N estimations from all N tomo-scans. Please see Supplementary Information for visualization. More details on the gradient calculation can be found in refs. ^{40,42}.

The data were pre-processed with multi-slice ptychography in order to properly address the sample at larger angles, where 2D ptychography may not be suitable since the effective optical path length becomes thicker than the depth of focus of the probe. The depth of focus (DOF) of the probe is approximately $\frac{\lambda}{2(\text{NA})^2} = 2.82 \mu\text{m}$. When the sample is rotated by the largest angle for tomographic scanning, i.e. 70.4° , the optical path length increases up to $\frac{3.92}{\cos(70.4^\circ)} = 11.7 \mu\text{m} \simeq 4.46 \text{ DOF}$, so 5 slices should be sufficient to address the rotated sample.

Machine learning framework

Our neural network architecture is based on a 3D U-net structure^{27,28} augmented with multi-head axial self-attention (“axial self-attention” in short)²⁹. The motivation behind self-attention is similar to that of atrous convolution (or dilated convolution) in convolutional neural networks^{51–53}, which also seek to enhance the receptive field and account for multi-scale features. We decided to incorporate self-attention into our architecture instead of dilated convolution, as it generates attention maps that provide a deeper understanding of the neural network’s functions. In our network, the U-net directly transfers multi-scale encoded features to its decoder arm to preserve spatial information, and the axial attention augments the features with its global-range self-interactions.

The U-net backbone encoder design was influenced by the well-established architecture ResNet50⁵⁴ with some modifications so that it can accommodate 3D instead of 2D data. The architecture’s decoder then upsamples the features to result in isotropic voxels of linear size 14 nm. More details can be found in Supplementary Information.

The encoder’s low-dimensional manifolds are further enhanced by the axial self-attention which was proposed in order to reduce the computational complexity of multi-head self-attention (“self-attention” in short)³⁰. The axial self-attention factorizes 3D self-attention into three 1D axial self-attention modules, thus reducing the complexity from $O(N^3)$ to $O(3N)$. Each axial self-attention attends to voxels along one of x, y, z axes. Figure 7 visualizes learned attention weights p_{ij} that quantifies normalized “contribution” of other layers s_j ($j = 1, \dots, N$) to the layer s_i . We assume that the information of layer s_i is spread along the layers s_j ($j = 1, \dots, N$) due to lack of spatial resolution; therefore, the axial self-attention gathers the scattered information from the layers to resolve layer s_i with global-range



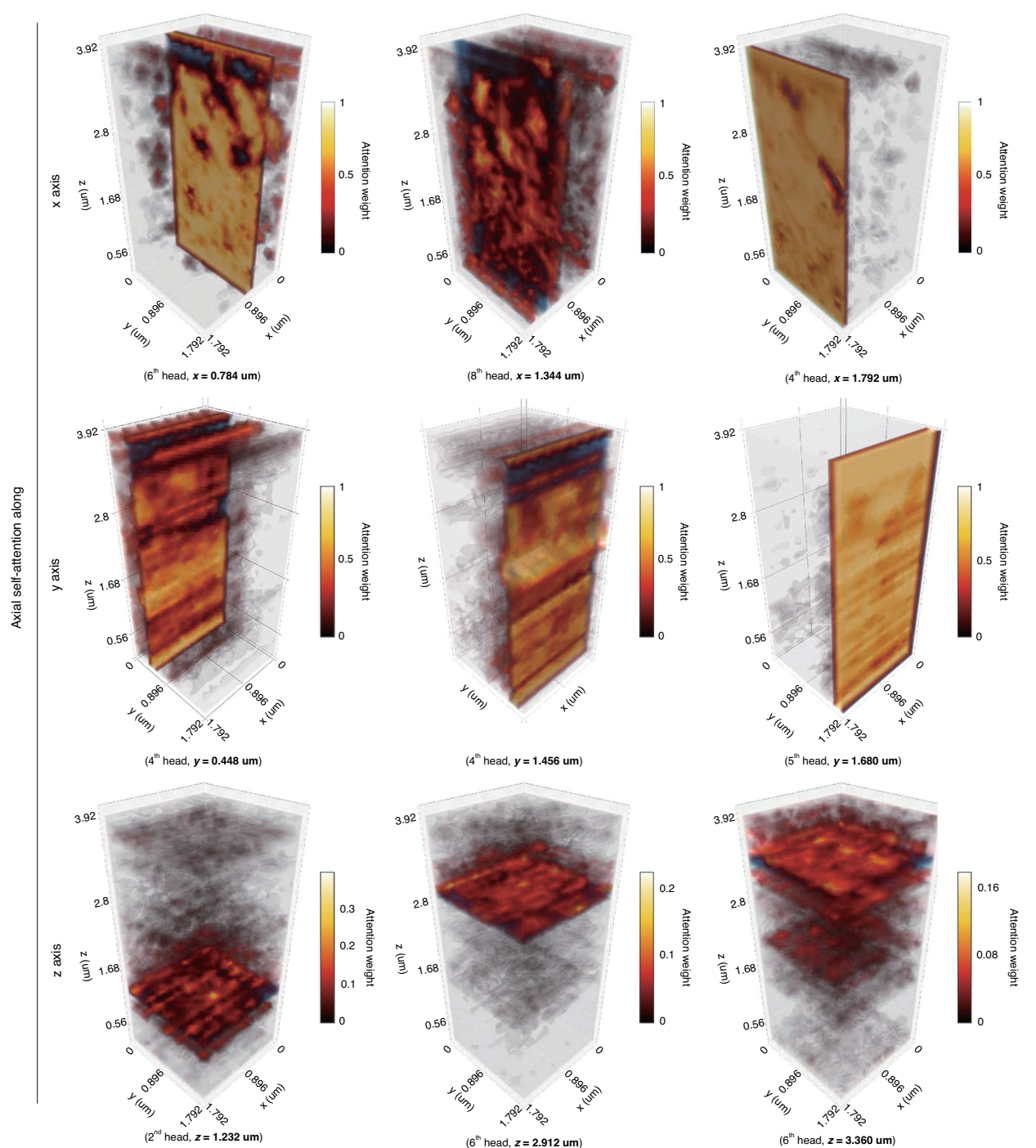


Fig. 7 Learned attention weight visualization. Learned attention weights of multi-head axial self-attentions along each x, y, z axis. Parentheses contain information on the selected attention head and the position of the layer of interest (blue, s_j) that attends to all layers (s_j ($j = 1, 2, \dots, N$)) with attention weights (red, α_{ij}^k), showing importance of s_j to s_i

interactions. Note that in this paper, we used Pytorch instead of the original Tensorflow implementation²⁹, and our code should be publicly available in <https://github.com/iksungk/APT>.

Training and testing environments

To prepare a paired dataset for training and testing, both of the Approximant and the gold standard are divided into smaller volumes of $1.792 \times 1.792 \times 3.92 \mu\text{m}^3$



with 50% lateral overlap. Then, we split the paired dataset into two non-overlapping sub-datasets. One set is reserved for training, and the other for testing. The training and test samples were drawn so as to not be correlated accidentally by spatial overlap during the ptycho- and tomo-scan operations. (In Supplementary Information, we provide comprehensive details on how the datasets were partitioned, along with additional evidence demonstrating the robustness of our algorithm to different partitioning methods.) To mitigate the generalization problem that arises when pre-trained networks fail to perform well on testing data that significantly departs from the training data, we trained the network on a portion of the IC and applied the trained network to the remainder. Transfer learning can be utilized to reduce the amount of training required for the network when working with new ICs that have different design rules.

For training, we use negative Pearson Correlation coefficient (NPCC) as the training loss function^{31,32,55} and the Adam optimizer for stochastic gradient descent optimization⁵⁶ with initial learning rate of 2×10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, and without weight decay. We also update the learning rate schedule according to a polynomial rule⁵⁷ as

$$lr(\text{epoch}) = lr(0) \times \left(1 - \frac{\text{epoch}}{T}\right)^p \quad (9)$$

where $lr(0) = 2 \times 10^{-4}$. We followed the convention of Ref. ⁵⁸ by setting $p = 0.9$. Additionally, we aimed to reduce the final learning rate to 1/100 of the initial rate, which was accomplished by setting $T = 200$. We trained the model for 150 epochs and utilized a mini-batch learning strategy⁵⁹ with a batch size of 4 to stabilize the process. This batch size was chosen as it was optimal for the computational memory limitations of our machine. Upon completion of the training process, the network is loaded and fixed with the optimal weights, and used to reconstruct the test volume ($4.48 \times 93.2 \times 3.92 \mu\text{m}^3$), as shown in Figs. 2, 4, and 6.

For all computational procedures, i.e. pre-processing and training and testing processes, we used the MIT Supercloud with a Intel Xeon Gold 6248 CPU with 384 GB RAM and dual NVIDIA Volta V100 GPUs with 32 GB VRAM. Once the network was trained, it took 45 s to generate the reconstruction over the test volume.

Quantitative metrics

Because each voxel on an IC is occupied by a single material, even if ICs are printed with various materials such as copper, aluminum, and tungsten, ICs can be comfortably classified into M -ary labels irrespective of the printing material. To further simplify, we binarize the gold standard by thresholding according to the presence of a

metal or silicon within each voxel. The gold standard reconstruction, however, may still be ambiguous especially for longitudinal features due to the missing wedge in the Fourier domain as it still does not cover the entire angular range, i.e. $\pm 90^\circ$, due to the tomographic scheme. Since the gold standard also suffers from extensive errors in these ambiguous layers, we exclude them from our quantitative evaluations as well. More details can be found in Supplementary Information.

The quantitative comparisons in Figs. 3 and 5 use four different quantitative metrics to illustrate different aspects of the reconstructions. The first two are correlative metrics: the PCC⁵⁵ and the MS-SSIM with the same weights as in the original reference³⁴.

The remaining two metrics are the DSC³⁵ and the BER. The former is a widely accepted similarity measure in image segmentation to compare an algorithm output against its reference in medical applications^{60,61}. The BER measures the ratio of erroneously classified voxels over the total voxels, and it is allowable because of our binarization approach. Both of these metrics are probabilistic in the sense that they involve the estimation of probability density functions. They are obtained as

$$\text{DSC} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FN} + \text{FP}} \quad (10)$$

$$\text{BER} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (11)$$

where TP, TN, FP, and FN indicate the number of true positives, true negatives, false positives, and false negatives, respectively. For the gold standard, the binary thresholds and prior probabilities $p(0), p(1)$ required for these quantities were estimated by an Expectation Maximization (EM) procedure. For the tests, we used Bayes' rule $p(x|0)p(0) = p(x|1)p(1)$ with $p(0), p(1)$ same as for the gold standard.

Furthermore, we have conducted a resolution analysis of the reconstructions shown in Figs. 3 and 5 using the Fourier ring coefficient⁶², which is detailed in the Supplementary Information.

Acknowledgements

We are grateful to Jung Ki Song, Mo Deng, Baoliang Ge, William Harrod, Ed Cole, Zachary Levine, Bradley Alpert, Nina Weiss-Bernstein, Lee Oesterling, and Antonio Orozco for helpful discussions and comments. Funding from the Intelligence Advanced Research Projects Activity, Office of the Director of National Intelligence (IARPA-ODNI), contract FA8650-17-C-9113 is gratefully acknowledged. The MIT SuperCloud and Lincoln Laboratory Supercomputing Center provided resources (high performance computing, database, consultation) that have contributed to the research results reported within this paper. I. Kang acknowledges support from Korea Foundation for Advanced Studies (KFAS). This research used resources of the Advanced Photon Source, a U.S. Department of Energy (DOE) Office of Science User Facility, operated for the DOE Office of Science by Argonne National Laboratory under Contract No. DE-AC02-06CH11357. The views and conclusions contained herein are those of



the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODN, IARPA, or the US Government.

Author details

¹Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ²Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ³Argonne National Laboratory, Lemont, IL 60439, USA. ⁴Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 CREATE Way, Singapore 138602, Singapore. ⁵Present address: Department of Molecular and Cell Biology, University of California, Berkeley, CA 94720, USA. ⁶Present address: Singapore-MIT Alliance for Research and Technology (SMART) Centre, 1 CREATE Way, Singapore 138602, Singapore. ⁷Present address: Center for Transformative Science, ShanghaiTech University, 201210 Shanghai, China

Author contributions

I.K., Z.W., S.V., and G.B. conceived the research; I.K. developed the machine learning algorithm and performed image processing and data analysis; I.K. and Z.W. designed the derivation of the Approximant and the neural network training process; Y.J., Y.Y., J.D., and J.K. conducted the experiments; and all authors contributed to the preparation of the paper and the discussion.

Conflict of interest

The authors declare no competing interests.

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41377-023-01181-8>.

Received: 16 January 2023 Revised: 9 May 2023 Accepted: 10 May 2023
Published online: 30 May 2023

References

- Hoppe, W. Beugung im inhomogenen primärstrahlwellenfeld. I. Prinzip einer phasenmessung von elektronenbeugungsinterferenzen. *Acta Crystallogr. Sect. A* **25**, 495–501 (1969).
- Hegerl, R. & Hoppe, W. Dynamische theorie der kristallstrukturanalyse durch elektronenbeugung im inhomogenen primärstrahlwellenfeld. *Ber. Bunsenges. Phys. Chem.* **74**, 1148–1154 (1970).
- Nellist, P. D., McCallum, B. C. & Rodenburg, J. M. Resolution beyond the 'information limit' in transmission electron microscopy. *Nature* **374**, 630–632 (1995).
- Gerchberg, R. W. A practical algorithm for the determination of phase from image and diffraction plane pictures. *Optik* **35**, 237–246 (1972).
- Faulkner, H. M. L. & Rodenburg, J. M. Movable aperture lensless transmission microscopy: a novel phase retrieval algorithm. *Phys. Rev. Lett.* **93**, 023903 (2004).
- Rodenburg, J. M. & Faulkner, H. M. L. A phase retrieval algorithm for shifting illumination. *Appl. Phys. Lett.* **85**, 4795–4797 (2004).
- Thibault, P. et al. High-resolution scanning X-ray diffraction microscopy. *Science* **321**, 379–382 (2008).
- Thibault, P. & Guizar-Sicairos, M. Maximum-likelihood refinement for coherent diffractive imaging. *New J. Phys.* **14**, 063004 (2012).
- Radon, J. On the determination of functions from their integrals along certain manifolds. *MathematischePhysikalische* **69**, 262–277 (1917).
- Radon, J. On the determination of functions from their integral values along certain manifolds. *IEEE Trans. Med. Imag.* **5**, 170–176 (1986).
- Bracewell, R. N. Strip integration in radio astronomy. *Aust. J. Phys.* **9**, 198–217 (1956).
- Bracewell, R. N. & Riddle, A. C. Inversion of fan-beam scans in radio astronomy. *Astrophys. J.* **150**, 427 (1967).
- Gordon, R., Bender, R. & Herman, G. T. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography. *J. Theor. Biol.* **29**, 471–481 (1970).
- Gilbert, P. Iterative methods for the three-dimensional reconstruction of an object from projections. *J. Theor. Biol.* **36**, 105–117 (1972).
- Andersen, A. H. & Kak, A. C. Simultaneous algebraic reconstruction technique (SART): a superior implementation of the art algorithm. *Ultrason. Imaging* **6**, 81–94 (1984).
- Bouman, C. & Sauer, K. A generalized Gaussian image model for edge-preserving map estimation. *IEEE Trans. Image Process.* **2**, 296–310 (1993).
- Dierolf, M. et al. Ptychographic X-ray computed tomography at the nanoscale. *Nature* **467**, 436–439 (2010).
- Holler, M. et al. High-resolution non-destructive three-dimensional imaging of integrated circuits. *Nature* **543**, 402–406 (2017).
- Deng, J. J. et al. The velociprobe: an ultrafast hard X-ray nanoprobe for high-resolution ptychographic imaging. *Rev. Sci. Instrum.* **90**, 083701 (2019).
- Huang, X. J. et al. Fly-scan ptychography. *Sci. Rep.* **5**, 9074 (2015).
- Lempitsky, V., Vedaldi, A. & Ulyanov, D. Deep image prior. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. p. 9446–9454 (IEEE, 2018).
- Mataev, G., Elad, M. & Milanfar, P. DeepRED: deep image prior powered by RED. *arXiv* <https://doi.org/10.48550/arXiv.1903.10176> (2019).
- Kappeler, A. et al. Ptychnet: CNN based Fourier ptychography. *2017 IEEE International Conference on Image Processing (ICIP)*. p. 1712–1716 (IEEE, 2017).
- Nguyen, T. et al. Deep learning approach for Fourier ptychography microscopy. *Opt. Express* **26**, 26470–26484 (2018).
- Cherukara, M. J. et al. AI-enabled high-resolution scanning coherent diffraction imaging. *Appl. Phys. Lett.* **117**, 044103 (2020).
- Goy, A. et al. Low photon count phase retrieval using deep learning. *Phys. Rev. Lett.* **121**, 243902 (2018).
- Ronneberger, O., Fischer, P. & Brox, T. U-net: convolutional networks for biomedical image segmentation. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. p. 234–241 (Springer, 2015).
- Çiçek, Ö. et al. 3D U-net: learning dense volumetric segmentation from sparse annotation. *18th International Conference on Medical Image Computing and Computer-Assisted Intervention*. p. 424–432 (Springer, 2016).
- Wang, H. Y. et al. Axial-DeepLab: stand-alone axial-attention for panoptic segmentation. *16th European Conference on Computer Vision*. p. 108–126 (Springer, 2020).
- Vaswani, A. et al. Attention is all you need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*. p. 6000–6010 (Curran Associates Inc., 2017).
- Goy, A. et al. High-resolution limited-angle phase tomography of dense layered objects using deep neural networks. *Proc. Natl Acad. Sci. USA* **116**, 19848–19856 (2019).
- Kang, I., Goy, A. & Barbastathis, G. Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views. *Light Sci. Appl.* **10**, 74 (2021).
- Benesty, J. et al. Pearson correlation coefficient. in *Noise Reduction in Speech Processing* (eds Benesty, J. et al) p. 1–4 (Springer, 2009).
- Wang, Z., Simoncelli, E. P. & Bovik, A. C. Multiscale structural similarity for image quality assessment. *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, 2003. p. 1398–1402 (IEEE, 2003).
- Dice, L. R. Measures of the amount of ecologic association between species. *Ecology* **26**, 297–302 (1945).
- Pelt, D. M. et al. Integration of tomopy and the ASTRA toolbox for advanced processing and reconstruction of tomographic synchrotron data. *J. Synchrotron Radiat.* **23**, 842–849 (2016).
- Chambolle, A. & Pock, T. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vis.* **40**, 120–145 (2011).
- Wolf, E. New theory of partial coherence in the space–frequency domain. Part I: spectra and cross spectra of steady-state sources. *J. Opt. Soc. Am. A* **72**, 343–351 (1982).
- Thibault, P. & Menzel, A. Reconstructing state mixtures from diffraction measurements. *Nature* **494**, 68–71 (2013).
- Tsai, E. H. R. et al. X-ray ptychography with extended depth of field. *Opt. Express* **24**, 29089–29108 (2016).
- Suzuki, A. et al. High-resolution multislice X-ray ptychography of extended thick objects. *Phys. Rev. Lett.* **112**, 053903 (2014).
- Chen, Z. et al. Electron ptychography achieves atomic-resolution limits set by lattice vibrations. *Science* **372**, 826–831 (2021).
- Baguer, D. O., Leuschner, J. & Schmidt, M. Computed tomography reconstruction using deep image prior and learned reconstruction methods. *Inverse Probl.* **36**, 094004 (2020).



44. Kang, I. et al. Accelerated deep self-supervised ptycho-laminography for three-dimensional nanoscale imaging of integrated circuits. *arXiv* <https://doi.org/10.48550/arXiv.2304.04597> (2023).
45. Helfen, L. et al. High-resolution three-dimensional imaging of flat objects by synchrotron-radiation computed laminography. *Appl. Phys. Lett.* **86**, 071915 (2005).
46. Holler, M. et al. Three-dimensional imaging of integrated circuits with macro-to nanoscale zoom. *Nat. Electron.* **2**, 464–470 (2019).
47. Odstrčil, M., Menzel, A. & Guizar-Sicairos, M. Iterative least-squares solver for generalized maximum-likelihood ptychography. *Opt. Express* **26**, 3108–3123 (2018).
48. Wakonig, K. et al. *Ptychoshelves*, a versatile high-level framework for high-performance analysis of ptychographic data. *J. Appl. Crystallogr.* **53**, 574–586 (2020).
49. Dong, C. et al. Learning a deep convolutional network for image super-resolution. *13th European Conference on Computer Vision*. p. 184–199 (Springer, 2014).
50. Dong, C. et al. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **38**, 295–307 (2016).
51. Pelt, D. M. & Sethian, J. A. A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl Acad. Sci. USA* **115**, 254–259 (2018).
52. Chen, L. C. et al. Rethinking atrous convolution for semantic image segmentation. *arXiv* <https://doi.org/10.48550/arXiv.1706.05587> (2017).
53. Wu, Z. L. et al. *Computational Optical Sensing and Imaging 2022* (Optica Publishing Group, 2022).
54. He, K. M. et al. Deep residual learning for image recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition*. p. 770–778 (IEEE, 2016).
55. Li, S. et al. Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).
56. Kingma, D. P. & Ba, J. Adam: a method for stochastic optimization. *3rd International Conference on Learning Representations*. (ICLR, 2015).
57. Cheng, B. W. et al. Panoptic-DeepLab: a simple, strong, and fast baseline for bottom-up panoptic segmentation. *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. p. 12472–12482 (IEEE, 2020).
58. Liu, W., Rabinovich, A. & Berg, A. C. ParseNet: looking wider to see better. *arXiv* <https://doi.org/10.48550/arXiv.1506.04579> (2015).
59. Li, M. et al. Efficient mini-batch training for stochastic optimization. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. p. 661–670 (ACM, 2014).
60. Milletari, F., Navab, N. & Ahmadi, S. A. V-net: fully convolutional neural networks for volumetric medical image segmentation. *2016 Fourth International Conference on 3D Vision (3DV)*. p. 565–571 (IEEE, 2016).
61. Zou, K. H. et al. Statistical validation of image segmentation quality based on a spatial overlap index¹: scientific reports. *Acad. Radiol.* **11**, 178–189 (2004).
62. Koho, S. et al. Fourier ring correlation simplifies image restoration in fluorescence microscopy. *Nat. Commun.* **10**, 3103 (2019).

