# High-resolution limited-angle phase tomography of dense layered objects using deep neural networks

Alexandre Goy[a,1], Girish Rughoobur[b], Shuai Li[a], Kwabena Arthur[a], Akintunde I. Akinwande[b], and George Barbastathis[a,c]

[a]3D Optics Laboratory, Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Microsystems Technology Laboratories, Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139; and [c]BioSystems and bioMechanics (BioSyM) Interdisciplinary Research Group, Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 117543, Singapore

Edited by Terrence J. Sejnowski, Salk Institute for Biological Studies, La Jolla, CA, and approved August 19, 2019 (received for review December 14, 2018)

**We present a machine learning-based method for tomographic reconstruction of dense layered objects, with range of projection angles limited to ±10°. Whereas previous approaches to phase tomography generally require 2 steps, first to retrieve phase projections from intensity projections and then to perform tomographic reconstruction on the retrieved phase projections, in our work a physics-informed preprocessor followed by a deep neural network (DNN) conduct the 3-dimensional reconstruction directly from the intensity projections. We demonstrate this single-step method experimentally in the visible optical domain on a scaled-up integrated circuit phantom. We show that even under conditions of highly attenuated photon fluxes a DNN trained only on synthetic data can be used to successfully reconstruct physical samples disjoint from the synthetic training set. Thus, the need for producing a large number of physical examples for training is ameliorated. The method is generally applicable to tomography with electromagnetic or other types of radiation at all bands.**

deep learning | tomography | imaging through scattering media

Tomography is the quintessential inverse problem. Since the interior of a 3-dimensional (3D) object is not accessible noninvasively, the original insight of tomographic approaches was to illuminate through from multiple angles of incidence and then process the resulting projections to reconstruct the interior slice by slice (1–3). In the simplest case, when diffraction is negligible and the illumination is collimated, as is generally permissible to assume for X-ray attenuation (4–7) and electron scattering (8–10) in the far field and for features of size ∼1 μm and above, the object's interior is represented by its Radon transform (11) of line integrals along straight parallel paths. The interior of the volume is then reconstructed by use of the Fourier-slice theorem for the Radon projections. On the other hand, if the X-ray beam is not collimated but spherical, then the slice-by-slice approach is no longer applicable and full volumetric reconstruction is required (12, 13). Even when the object is available for observation from the full 360° range of projection angles, these instances of tomography are all highly ill-posed because the Fourier-slice property results in uneven coverage of the Fourier space with the high spatial frequencies ending up underrepresented. Ill-posedness increases when the angular range is limited because then an entire cone of spatial frequencies goes missing from the measurement. Alternatively, in this case, tomosynthesis (14) utilizes sheared (rather than rotated) projections to bring slices from within the interior into focus, but with lower contrast since emission from the rest of the volume remains as background.

Additional challenges occur when the inverse problem of interest is to reconstruct in 3D the index of refraction, rather than the attenuation. If the object features are large enough compared to the wavelength, such that diffraction may still be neglected, and the index variations through the object volume are relatively small, then each projection may be modeled as a set of Fermat integrals of phase delay along approximately straight lines. The phase integrals may be obtained, for example, using holographic interferometry (15, 16) or transport of intensity (17). For smaller-sized features and still assuming weak scattering (first-order Born approximation), the projection integrals are instead obtained along curved paths on the surface of the Ewald sphere, a method referred to as diffraction tomography (18, 19). By decoupling the problem into 2 parts, first phase projection retrieval, followed by tomography, these approaches enjoy the benefit of using the advanced algorithms in the 2 respective research fields. However, there is also the danger that errors generated independently during each step may amplify each other. Finally, when strong scattering may no longer be neglected, all 2-step approaches become questionable because the interpretation of the first step as line integrals is no longer valid.

Generally, ill-posed inverse problems are solved by regularized optimization. If $f$ is the object and $g$ the measurement, then the object estimate $\hat{f}$ is obtained as (20, 21)

$$\hat{f} = \underset{f}{\operatorname{argmin}} \left\{ \|Hf - g\|^2 + \alpha\Phi(f) \right\}. \quad [1]$$

Here, $H$ is the forward operator relating the measurement to the object, $\Phi$ is the regularizer expressing prior knowledge about the object, and $\alpha$ is the regularization parameter controlling the competition between the 2 terms. The prior may be thought of as rejecting solutions to the inverse problem that are known to violate known properties of the class of objects being imaged; for example, if the class where $f$ belongs is known to have sharp edges, then the regularizer should be applying a high penalty to blurry solutions $\hat{f}$. Thus, the inherent uncertainty due to ill-posedness is reduced. Sparsity-promoting compressive priors (22–25) found some of their first successes in tomographic reconstruction (26, 27). Compressive sensing is directly implemented through a proximal gradient solution to Eq. **1** if a set of basis

## Significance

**We demonstrate that it is possible to use deep neural networks to produce tomographic reconstructions of dense layered objects with small illumination angle as low as 10°. It is also shown that a DNN trained on synthetic data can generalize well to and produce reconstructions from experimental measurements. This work has application in the field of X-ray tomography for the inspection of integrated circuits and other materials studies.**

Author contributions: A.I.A. and G.B. designed research; A.G. performed research; G.R., S.L., and K.A. contributed new reagents/analytic tools; A.G. analyzed data; and A.G., G.R., and G.B. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.



[1]To whom correspondence may be addressed. Email: agoy@goyman.com.

First published September 16, 2019.

functions where the object class is sparse is a priori known. Alternatively, if a database of representative objects is available, then these examples may be used to learn the optimal set of basis functions as a dictionary (28, 29).

Rapid recent developments in the field of machine learning, and deep neural networks (30) (DNNs) in particular, have provided an additional set of tools and insights for inverse problems. It may be shown (31, 32) that recurrent or unfolded multistage DNN architectures are formally equivalent to the iterative solution to the inverse problem in Eq. **1** where the prior $\Phi$ need no longer be known or depend on sparsity; instead, examples guide the discovery of the prior through the DNN training process. Simpler learning architectures, where $g$ is fed to the DNN directly or after first passing through a preprocessor, have been used for retrieval of phase from intensity (33–36); 3D holographic reconstruction (37–39); superresolution photography (40–42) and microscopy (43); imaging through scatter (44–47); and imaging under extremely low light conditions in the 3 contexts computational ghost imaging (48), consumer-camera photography (49), and phase retrieval (50).

Multistage DNN architectures have been shown to yield high-quality reconstructions in numerous Radon tomography configurations (32, 51–56). Recently, Nguyen et al. (57) used the inverse Radon transform for optical tomography with a single-stage DNN intended to partially correct for the assumption of line integrals breaking down.

In this paper, we apply a Fourier-based beam propagation method (BPM) (58) as a preprocessing step immune from any Radon assumptions. The strongly scattering object is illuminated by a parallel beam under a limited angular range of $20^\circ$, i.e., $\pm 10^\circ$ from the reference axis. Unlike the earlier works on refractive index tomography referenced above, we do not perform a phase retrieval step; rather, the intensity measurements are preprocessed to produce directly an initial crude 3D guess of the object's interior. This crude guess is then fed to our machine-learning algorithm. The preprocessing step is necessary because, even if we did convert intensity to phase, the results would not be interpretable as line integrals under our experimental conditions. Moreover, by merging phase retrieval and tomography into a single step, our algorithm becomes less sensitive to error accrual.

Large datasets, typically consisting of more than 5,000 examples, are generally required for DNN training. That is feasible in many cases through spatial light modulators (33, 57). However, in many cases of interest spatial light modulators have insufficient space-bandwidth product or are unavailable (e.g., in X-rays); and alternatives to generate physical specimens are expensive or restricted due to proprietary processes. Instead, our approach is to train the DNN on purely synthetic data with the rigorous BPM forward model and then use a physical test specimen (phantom) to test the reconstruction quality with well-calibrated ground truth in experiments.

We chose to design our phantom as emulating the 3D geometry of integrated circuits (ICs). These would normally be inspected with X-rays, so we scaled up the feature dimensions in the phantom for visible wavelengths. The advantage of this choice is that IC layouts provide strong geometrical priors, e.g., Manhattan geometries, and our phantom also exhibited large spatial gradients and refractive index contrast to strengthen scattering. Thus, our methodology is directly applicable to all cases of tomography at optical wavelengths, e.g., 3D-printed specimen characterization and identification and biological studies in cells and tissue with moderate scattering properties. In each case, testing the ground truth would require the fabrication or the accurate simulation of different phantoms meeting the corresponding priors.

There is also value in the study of emulating X-ray inspection of ICs at visible wavelengths, as extensive outsourcing by the IC industry has created a growing concern that the ICs delivered to the customer may not match the expected design and that malicious features may have been added (59). However, in our emulation the phase contrast of the features against the background and the Fresnel number are both higher than typical corresponding IC configurations even at soft X-ray wavelengths.

One advantage of deep learning for inverse problems is speed. Solving **[1]** separately for each instance of $g$ is computationally intensive, and training a DNN is even more so. This is because both operations are iterative, and the latter is run on large datasets. On the other hand, once the DNN has learned the inverse map from the preprocessed version of $g$ to $\hat{f}$, the computations are feedforward only. For example, the IC layout priors we exploit here could, in principle, also be learned by dictionaries—but, under strong scattering conditions, the latter would require iterative optimization of Eq. **1** with the forward operator $H$ itself consisting of an expensive computational procedure in each iteration. In our approach, the preprocessing performed prior to the DNN is the most time-consuming operation, and therefore we aim at simplifying the preprocessing step as much as we can, i.e., tolerate a crude approximation, and leave it to the DNN to correct it. In our case, the execution time is 51 s (out of which only 300 ms are taken by the DNN, with the rest being the preprocessor), while learning tomography (60, 61), which is based on a similar gradient descent algorithm, takes 212 s and yields inferior reconstructions (see Fig. 4 for this, and see *Materials and Methods* for the computing hardware implementation).

## Optical Experiment

We prepared a series of 4 glass wafers with etched structures representing patterned layers from an actual IC design. A schematic cross-section of the sample is shown in Fig. 1*A*. The glass plates are held together and aligned in a custom-made holder. Immersion oil is added between the plates to minimize parasitic reflections and also tune the phase shift associated with the patterns. The pattern depth was measured to be 575 nm, yielding a phase shift of $-0.33$ rad for the particular oil used. Note that the phase shift is negative as the refractive index of the oil is lower than the refractive index of the glass. Details about the sample preparation and phase-shift measurements are given in *Materials and Methods*. The particular patterns etched on the sample are shown in Fig. 1*B*.

The experimental apparatus is detailed in Fig. 2. A collimated monochromatic plane wave from a continuous wave (CW) laser is incident on the sample, which is mounted on a 2-axis rotation stage. The sample is imaged through a demagnifying telescope to increase the field of view. The detector (an EM-CCD camera) is defocused from the image plane to simulate free-space propagation in an X-ray experiment where no imaging system can be used. Further details are given in *Materials and Methods*.

The strength of diffraction effects can be quantified with the Fresnel number $F = a^2/(\lambda d)$, where $\lambda$ is the wavelength, $d$ the propagation distance, and $a$ the characteristic feature size of the object. The smaller the Fresnel number is, the stronger the effects of diffraction. For the glass phantom considered here and a defocus of 58 mm, $F = 0.7$ for the smallest features and $F = 5.5$ for the largest. The diffraction pattern is digitized on the camera for different sample orientations. This series of measurements is then passed through a numerical algorithm, described in the next section, whose aim is to yield a first approximate reconstruction, hereafter referred to as the "approximant."

## Computation of the Approximant

As mentioned above, the task of the DNN is significantly facilitated if the raw measurements are preprocessed to give an approximation of the solution. We use a simple gradient descent

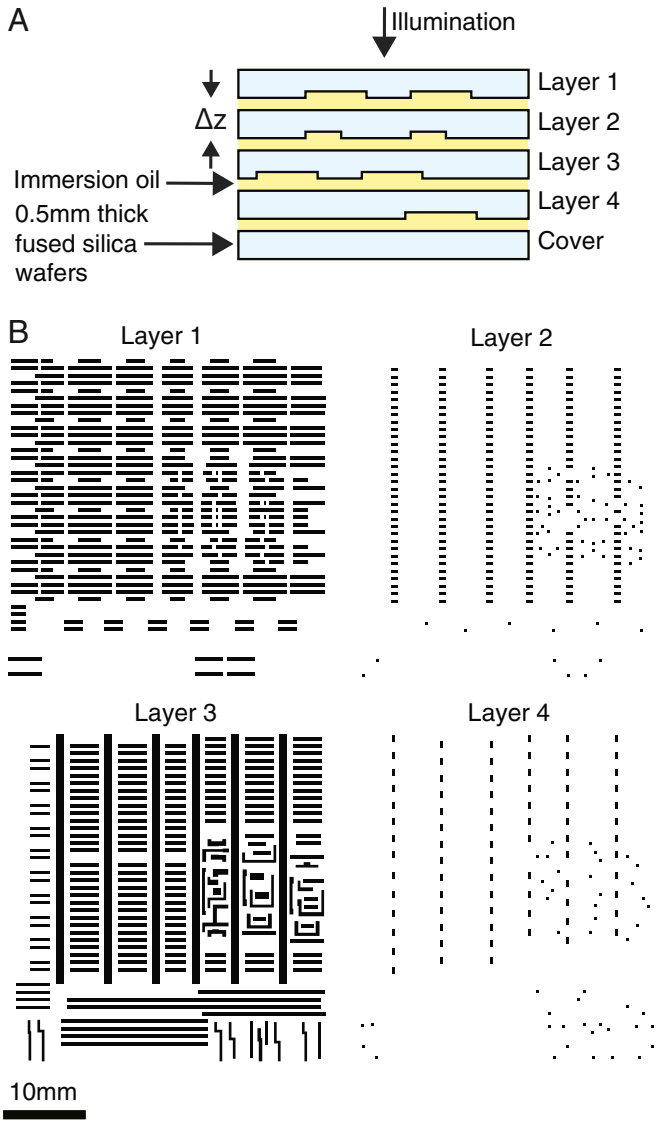


**Fig. 1.** (*A*) Sample cross-section. The depth of the etched patterns was measured (*Materials and Methods*) and the refractive index of the oil was controlled to achieve a known phase shift of −0.32 rad. $\Delta z = 0.5$ mm. (*B*) IC patterns used for each of the 4 layers. The white background represents the original wafer thickness and the black areas indicate where the wafer has been etched.

method to generate the approximant of the sample refractive index distribution.

Light propagation through the object can be computed using the following split-step Fourier BPM. In this method each sample layer $l$ is modeled as a 2D complex mask $f_l(x,y) = \exp[\alpha_l(x,y) + j\varphi_l(x,y)]$ and the space between 2 successive layers as Fresnel propagation through a homogeneous medium whose index of refraction equals the average refractive index of the sample, as

$$u_l = \mathcal{F}^{-1}\left\{\mathcal{F}\{u_{l-1}f_{l-1}\}(k_x,k_y)\,e^{-j\left(k-\sqrt{k^2-k_x^2-k_y^2}\right)\Delta z}\right\}. \quad [2]$$

Here, $u_l(x,y)$ is the optical field at layer $l$, $\mathcal{F}$ the Fourier transform, $\Delta z$ the distance between layers, and $k$ the wavenumber in the medium between layers. Each measurement is a collection of $N_v$ intensity patterns $g_i(x,y)$, with $i = 1,\ldots,N_v$, captured on the detector for each orientation $i$ of the sample. In this work, we assume that the sample is a pure phase object; i.e., $\alpha(x,y) = 0$. This assumption is valid for the glass phantom in the optical domain as well as for a sample composed of copper and silicon in the X-ray domain at 17 keV, where the phase contrast (real part of the refractive index contrast) is about 10 times larger than the absorption contrast (imaginary part of the refractive index contrast).

From the measurements, we produce an approximation $\tilde{\mathbf{f}}$ of the phase pattern $\varphi_l(x,y)$ for each layer $l$ in the sample. We use the steepest gradient descent method with a fixed number of iterations $K = 8$ to generate the approximant. In what follows, we represent the measurements (consisting of $M$ real pixel values) by a $(M \times 1)$ column vector $\mathbf{g}_i$ and the discretized object (consisting of $N$ real voxel values) by a $(N \times 1)$ column vector $\mathbf{f}$. We then define a cost function $J$ to minimize, consisting simply of a data fidelity term

$$J = \frac{1}{2}\sum_{i=1}^{N_v} \|H_i(\mathbf{f}) - \mathbf{g}_i\|_2^2, \quad [3]$$

where $H_i$ denotes the forward operator that maps the object function $\mathbf{f}$ to a prediction of the measurement $H_i(\mathbf{f})$, for a particular orientation $i$ of the sample. In the problem presented here, the optical field will first propagate through the sample $L$ layers, each of thickness $\Delta z$, and then in free space to the detector over a distance $d$. The forward operator can thus be written as a cascade of Fresnel propagation operations and thin mask multiplications corresponding to the object layers, i.e., successive applications of Eq. **2** written in operator form

$$H_i(\mathbf{f}) = |\mathbf{u}_{\text{det}}|^2 \quad [4]$$

$$\mathbf{u}_{\text{det}} = F_d \text{diag}[\mathbf{f}_L] \ldots F_{\Delta z}\text{diag}[\mathbf{f}_2]F_{\Delta z}\text{diag}[\mathbf{f}_1]\mathbf{u}_{\text{inc},i}, \quad [5]$$

where $\mathbf{f}_l$ is the vector of object function values in layer $l$, $F_{\Delta z}$ the Fresnel propagation operator over distance $\Delta z$, $\mathbf{u}_{\text{inc},i}$ the incident field, $\mathbf{u}_{\text{det}}$ the field on the detector, and $\text{diag}[\mathbf{v}]$ the diagonal matrix with vector $\mathbf{v}$ on the diagonal. The gradient descent iterative update can be written as

$$\mathbf{f}^{(k+1)} = \mathbf{f}^{(k)} - s(\nabla_{\mathbf{f}^{(k)}} J)^T, \quad [6]$$

where $\mathbf{f}^{(k)}$ is the object estimate at iteration $k$, $s$ the step size, and $\nabla_{\mathbf{f}^{(k)}} J$ the gradient of $J$ with respect to $\mathbf{f}$ evaluated at $\mathbf{f}^{(k)}$. We then set $\tilde{\mathbf{f}} = \mathbf{f}^{(K)}$ with $K$ chosen in advance, starting from $\mathbf{f}^{(0)} = \mathbf{0}$. The detailed derivation of the gradient for the particular

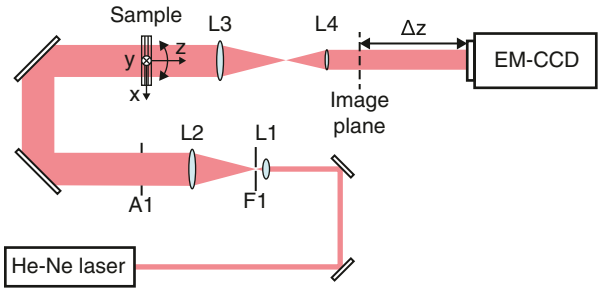


**Fig. 2.** Experimental apparatus: spatial filter and beam expander. L1 is 10×, 0.25 numerical aperture objective; L2 is a 100-mm lens, with a 5-μm pinhole F1 in the focal plane; L3 is a 200-mm lens; and L4 is a 100-mm lens. Aperture A1 cuts the outer diffraction lobes of the beam. The sample is mounted on a 2-axis rotation stage rotating along the *x* and *y* axes. The sample middle plane is imaged using a telescope lens system with magnification 0.50×. The camera is defocused by a distance $\Delta z = 58$ mm from the image plane.

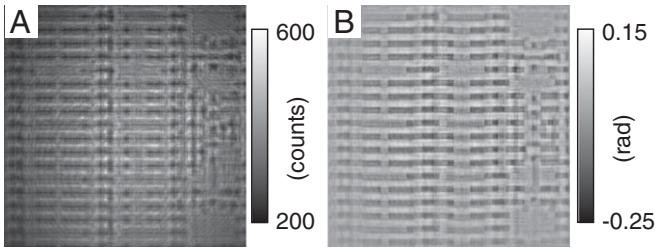


**Fig. 3.** (*A*) Examples of experimental intensity measurement for the sample orientation ($\theta_x = -10°, \theta_y = 0°$). (*B*) Phase approximant for IC layer 1 obtained from the collection of 22 intensity patterns at different orientations ($\theta_x = -10°, -8°, \ldots, +10°, \theta_y = 0°$) and ($\theta_x = 0°, \theta_y = -10°, -8°, \ldots, +10°$).

model in Eqs. **4** and **5** is given after the conclusion in the *Derivation of the Gradient* section. In Fig. 3, we give an example of 1 experimental intensity measurement ($\mathbf{g}_1$) taken from the series of tomographic projections and the corresponding approximant $\mathbf{f}^{(8)}$ obtained from the whole series using Eq. **6**.

## DNN Architecture and Training

We use a DNN to map the approximant to the final reconstruction $\hat{\mathbf{f}}$. The DNN is a convolutional neural network with a DenseNet architecture (62). The implementation is the same as the DNN used in ref. 45 except that the number of dense blocks was reduced to 3 in both the encoder and the decoder, as we empirically observed that using more dense blocks did not result in a significant improvement of the results. We produce a total of 5,500 synthetic sets of measurements obtained by simulating the optical apparatus with the beam propagation method in Eq. **2**. The synthetic measurements were subject to simulated shot noise and read noise equivalent to the noise levels found in the experiment. The shot noise was accounted for by converting the simulated measurement pixel intensities $I$ expressed in average photon count per pixel per integration period of the detector to integer numbers of photons $N$ following Poisson statistics. The actual optical power on the camera was measured with a power meter and converted to an average photon flux per detector pixel. Read noise following Gaussian statistics was subsequently added. The parameters (variance and average) of the noise were measured from a series of dark frames from the camera taken with the same gain (EM gain of 1) and integration time (2 ms) as the experimental measurements.

From each set of measurements, we produce a multilayer approximant using the gradient descent in Eq. **6**. The examples are split in a training set of 5,000 examples, a validation set of 450 examples, and a test set of 50 examples. Each set of measurements (1 example) comprises 22 views corresponding to different orientations of the sample. The DNN is then trained to map the approximant to the ground truth used for the simulation. Each layer of the sample is assigned to a different channel in the DNN. We use the negative Pearson correlation coefficient (NPCC $= -$PCC) as loss function and train in 20 epochs with a batch size of 16 examples. For 2 images $A$ and $B$, the PCC is defined as

$$\mathrm{PCC}(A,B) = \frac{\sum_i (A_i - \bar{A})(B_i - \bar{B})}{\sqrt{\sum_i (A_i - \bar{A})^2 \sum_i (B_i - \bar{B})^2}}, \qquad [7]$$

The PCC (and therefore the NPCC too) is agnostic to scale and offset; i.e., $\mathrm{PCC}(aA + b, cB + d) = \mathrm{PCC}(A, B)$ for $a, b, c, d \in \mathbb{R}$. As a consequence, the DNN, which is trained by minimizing the NPCC, may apply some offset and scaling to the reconstruction. These parameters are not easily predictable; however, for a given DNN they are constant once training is complete, which allows us to correct the reconstructions. Offset and scale are obtained by least-squares linear regression between the DNN output and the ground truth from the synthetic test set examples (not including the experimental example).

## Results

The method described in the previous sections was applied to the glass phantom shown in Fig. 1*B*. The synthetic measurements were subject to Poisson noise resulting from $10^3$ photon flux per detector pixel, equal to the experimental photon flux, and an additive Gaussian noise with a SD of 13 counts. For DNN training, we compared 2 sets of approximants, obtained with $K = 1$ and $K = 8$ with and without total variation (TV) regularization. In the case $K = 1$, the regularization parameter $\alpha$ was set to 0.1 (step size 0.1). We chose a smaller value of 0.04 for the case $K = 8$ (step size 0.05) because the proximal operator corresponding to the regularizer is applied at each iteration and its effect tends to accumulate. In the case $K = 8$, the particular choice for the number of iterations is an empirical trade-off between computation time and accuracy. The same optimization parameters (step size and number of iterations) were used to compute the approximant of the IC phantom, and the result for each layer is shown in Fig. 4 *E*–*H* for $K = 1$ and Fig. 4 *I*–*L* for $K = 8$. The DNN reconstruction results are summarized in Fig. 4 *M*–*P* ($K = 1$) and Fig. 4 *Q*–*T* ($K = 8$). The approximant and the DNN reconstruction represent the phase modulation imposed by each layer in the sample. The absolute phase carries no useful information; therefore we are free to offset the reconstructed phase by an arbitrary constant. In the DNN reconstructions in Fig. 4 *I*–*L*, the IC patterns (where the phase shift actually occurs) are typically reconstructed with 0 phase due to the rectified linear units (which project all negative values to 0) at the output layer of the DNN. We reassign the phase of the pattern to the nominal phase of $-0.33$ rad so that it can be visually compared to the ground truths in Fig. 4 *M*–*P*. An alternate approach leading to very similar results is to assign a 0 phase to the background.

The DNN reconstructions can be compared to those obtained using learning tomography (LT), a previously demonstrated optical tomography technique (60, 61) based on proximal optimization (FISTA) (63) with TV regularization (64). The role of the TV regularizer is to favor piecewise constant solutions while preserving sharp edges, which is especially well suited for IC patterns. LT was initially designed for holographic measurements and was modified here to work on intensity measurements by computing the gradient for the data fidelity term in Eq. **3**. The essential difference in the LT optimization is that a TV filter playing the role of a proximal operator is applied at each iteration on the current solution. The LT reconstructions for the experimental dataset are shown in Fig. 4 *A*–*D*. These particular reconstructions were obtained after 30 iterations of gradient descent, a step size of 0.05, a regularization parameter $\alpha = 0.04$, and 20 iterations for the TV regularizer at each step. The computation time of the $\mathbf{f}^{(8)}$ approximant is 51 s for $K = 8$ (no regularization) and 6 s for $K = 1$, including 570 ms for the DNN vs. 212 s for LT on the same processor (see *Materials and Methods* for hardware details).

In Table 1, we summarize the values of the PCC, which we use to quantify the quality of the reconstructions. The values are given for reconstructions on the synthetic test set (50 examples) and also the reconstruction of the single experimental example. Because the reconstruction quality turns out to depend strongly on the particular layer, we display the value for each of the 4 layers separately. As may be expected, the values for the LT are higher (better reconstructions) than those for the approximant as LT was run for 30 iterations vs. 8 for the approximant and that the latter was not regularized. The DNN reconstructions appear to be the best according to the PCC metric, which shows that, even while starting from a poor approximation, the DNN was

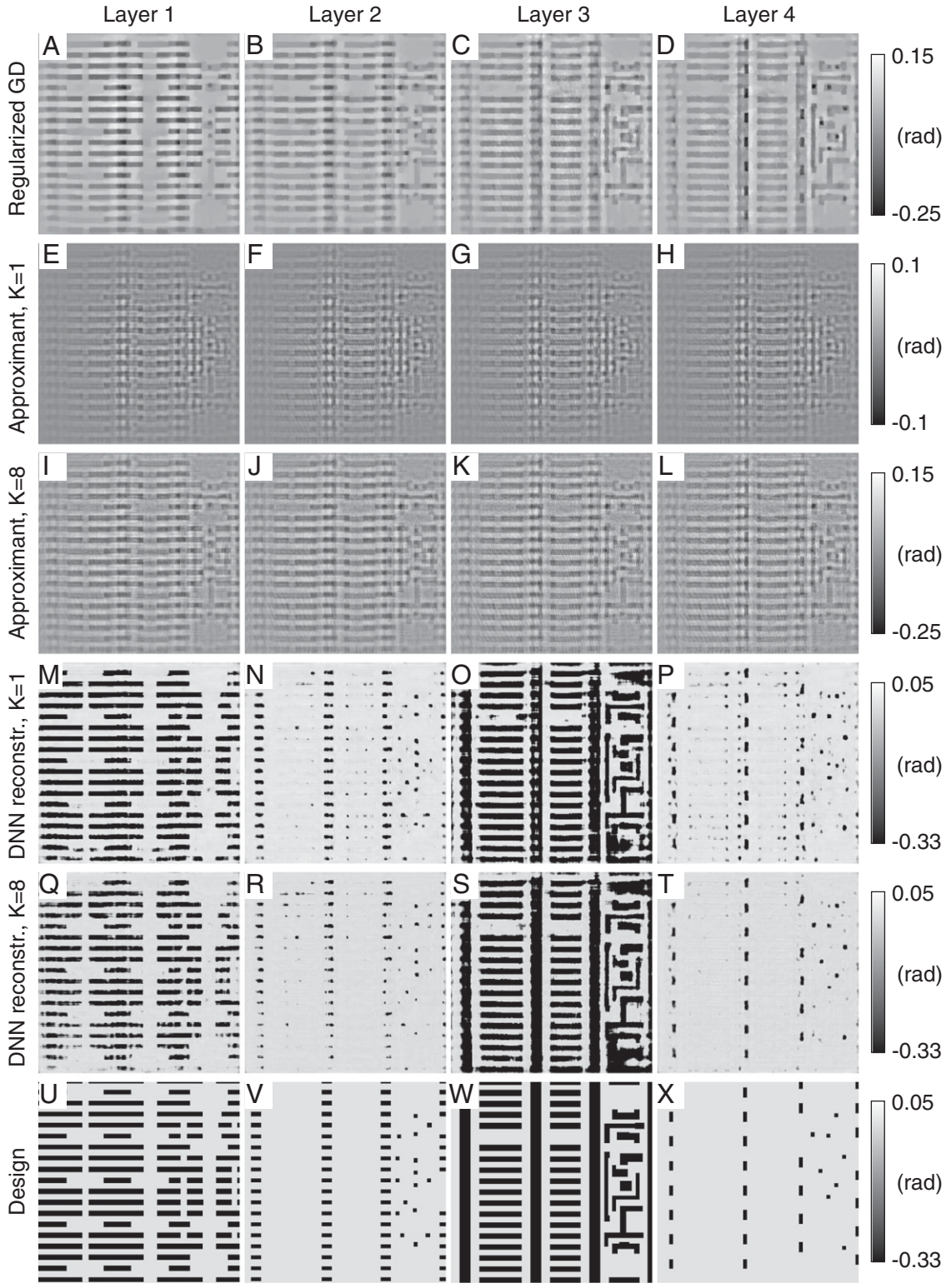


**Fig. 4.** (*A–D*) Proximal gradient descent with TV regularization, $K = 30$ iterations, for each layer 1 to 4. (*E–H*) Approximants generated from the experimental measurements with $K = 1$. (*I–L*) Approximants generated from the experimental measurements with $K = 8$. (*M–P*) Reconstructions from the DNN of each approximant *E–H*, respectively. (*Q–T*) Reconstructions from the DNN of each approximant *I–L*, respectively. (*U–X*) Idealized ground truth obtained from the sample specifications for layers 1 to 4. Note that the color bar range covers more than the range of the data, so there is no saturation effect on the images.

able to outperform LT. Note that a direct comparison between the performance of the DNN on the synthetic data and that on the experimental example is not fair because the ground truth is not known in the experiment. The ground truth used for the experimental example is an idealization from the design parameters used to fabricate the sample. We also indicate the values for reconstructions based on the regularized approximant (using the same regularization parameters as in the LT algorithm).

**Table 1. PCC, expressed in percentage (i.e., PCC × 100), of the reconstructions in the test set with respect to the ground truths for the approximant (not regularized) and the DNN reconstructions, labeled "DNN," obtained from the unregularized approximant**

| | | Approximant | | DNN | | DNN reg. | | LT | |
|---|---|---|---|---|---|---|---|---|---|
| K | Layer | Simul. | Exp. | Simul. | Exp. | Simul. | Exp. | Simul. | Exp. |
| 1 | 1 | 62 ± 7 | 48 | 99 ± 0.3 | 80 | 99 ± 0.4 | 72 | 91 ± 2 | 65 |
| 1 | 2 | 43 ± 5 | 22 | 97 ± 1 | 56 | 96 ± 1 | 45 | 79 ± 7 | 37 |
| 1 | 3 | 49 ± 9 | 41 | 99 ± 0.4 | 77 | 94 ± 5 | 76 | 89 ± 3 | 62 |
| 1 | 4 | 24 ± 7 | 7 | 95 ± 1 | 38 | 92 ± 2 | 42 | 76 ± 7 | 27 |
| 8 | 1 | 75 ± 63 | 63 | 100 ± 0.2 | 75 | 100 ± 0.1 | 76 | — | — |
| 8 | 2 | 57 ± 6.5 | 31 | 98 ± 0.7 | 44 | 99 ± 0.4 | 45 | — | — |
| 8 | 3 | 62 ± 6.5 | 52 | 99 ± 0.3 | 80 | 99 ± 0.3 | 79 | — | — |
| 8 | 4 | 41 ± 8.1 | 12 | 96 ± 0.8 | 48 | 98 ± 0.6 | 43 | — | — |

We show the 2 cases $K = 1$ and $K = 8$ for the approximant calculation. The LT solution is obtained with $K = 30$ and is indicated on the right. The values for the DNN trained with regularized approximants are labeled "DNN reg." The uncertainty values indicated correspond to the SD over the 50 examples of the test set. For each case, the values for the synthetic (simulated) and experimental examples are indicated in separated columns "Simul." and "Exp.," respectively. No uncertainty is given for the experimental case as it contains only 1 example.

In terms of PCC, there is no significant difference from the unregularized case.

The reconstructions based on regularized approximants are shown in Fig. 5. By comparing these images with the unregularized reconstructions shown in Fig. 4, and also by considering the value of the PCC in Table 1, we conclude that the regularization has little effect, especially on the experimental reconstructions. Moreover, the TV regularization may not operate as a favorable preconditioner for the DNN. The choice of the TV operator as a regularizer is arbitrary and only based on our assumption that the solution should be piecewise constant. In fact, because of the small angle range, the approximants for the different phantom layers are quite similar to each other and the regularization may cancel information that the DNN could use to discriminate between them. Layers 3 and 4 can be said to look visually better with the regularized approximant, but the situation is reversed for layers 1 and 2. Iterating more, i.e., using $K = 8$ vs. $K = 1$, yields slightly better results as can be expected intuitively, but the improvement is quite minute considering the increased computation time required to perform 7 more iterations.

In the regularized case $K = 1$ only we observed instability in the behavior of the DNN for the regularized approximants. For bipolar input (approximant containing both positive and negative values), one of the phantom layers (layer 3) would always be reconstructed to null values. As we are using a rectified linear unit as an activation function, this means that the output of one layer within the network displays only negative values. By offsetting the input to the DNN (approximant) so that all values are positive we were able to remove the problem (reconstructions of Fig. 5 *I–L*). For the regularized $K = 1$ case where this behavior was observed, the difference between approximant layers is the smallest; i.e., the failure may be due to the regularizer washing out the differences.

So far, we have reconstructed the phase shift distribution $\varphi(x, y)$ associated to each layer. In fact, it is possible, with the same method, to infer the refractive index $n(x, y)$ of the sample. For a given layer, the refractive index is simply given by $n(x, y) = \varphi(x, y)/(k\Delta z)$, where $\Delta z$ is the thickness of the layer. If the layer thicknesses are not known, one would instead slice the object into layers at finer spacing to meet the applicable Nyquist criterion.

## Conclusion

In this paper, we have demonstrated through an emulated X-ray experiment that DNNs can be used to improve the reconstruction of IC layouts from tomographic intensity measurements acquired within an angle range of $\pm 10^\circ$ along each lateral axis. The approximant obtained after 1 or several iterations of the steepest gradient method does not provide a reconstruction of sufficient quality for the purpose of IC integrity assessment. The DNN, however, exploits the strong prior contained in the object geometry and yields reconstructions that are significantly improved over the approximant. One of the main motivations for using DNN is indeed the speed of execution; therefore, we want to limit any unnecessary preprocessing. In fact, trying to improve the quality of the approximant by simply iterating more the gradient descent does not yield significant improvement.

One significant challenge for the method we demonstrated is to provide a proper training set for the DNN. In the case of ICs, the training set is simply given by the many real layouts that are available. For more generic objects, a problem needs first to be formulated to clearly define the class of object on which the training will be performed. This is, however, the case in general for problem solving that involves DNNs. What has been shown here is the compelling improvement that DNN can bring to a phase tomography problem when the class of object is known.

More generally, there is a trade-off between the specificity of the required priors and the "difficulty" of an inverse problem—measured as degree of ill-posedness, e.g., the ratio of largest to smallest eigenvalue in a linearized representation of the forward operator. The problem we addressed here is severely ill-posed due to the presence of strong scattering within the object and the limited range and number of angular measurements we allowed ourselves to collect. Therefore, the rather restricted nature of ICs as a class prior is justified; while, at the same time, our approach is addressing an indisputably important practical problem. Detailed determination of the relationship between the degree of ill-posedness and the complexity of the object class prior would be a worthwhile topic for future work.

**Derivation of the Gradient.** This derivation follows a path similar to the derivation given in ref. 61. We start from Eq. **3**:

$$J = \frac{1}{2}\sum_{i=1}^{N_v} \|H_i(\mathbf{f}) - \mathbf{g}_i\|_2^2 \qquad [8]$$

$$= \frac{1}{2}\sum_{i=1}^{N_v} \left( H_i(\mathbf{f})^T H_i(\mathbf{f}) - 2H_i(\mathbf{f})^T \mathbf{g}_i + \mathbf{g}_i^T \mathbf{g}_i \right). \qquad [9]$$

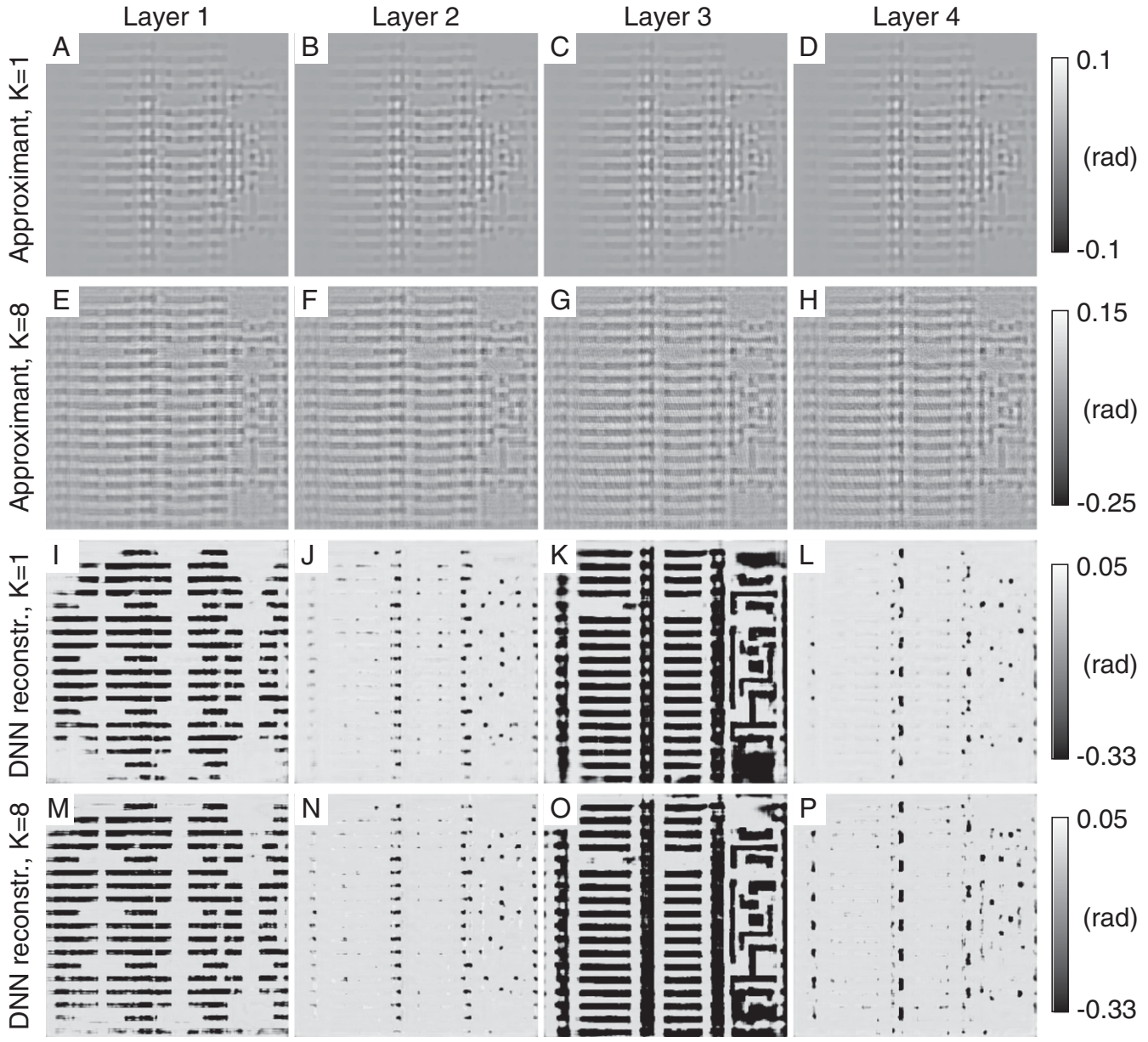


**Fig. 5.** (*A–D*) Approximants generated from the experimental measurements with $K = 1$ and TV regularization with $\alpha = 0.1$. (*E–H*) Approximants obtained with $K = 8$ and TV regularization with $\alpha = 0.04$. (*I–L*) Reconstructions from the DNN of each approximant *A–D*, respectively. (*M–P*) Reconstructions from the DNN of each approximant *E–H*, respectively.

The gradient of $J$ is defined as

$$\nabla_{\mathbf{f}} J = \left[ \frac{\partial J}{\partial \varphi_1}, \ldots, \frac{\partial J}{\partial \varphi_N}, \frac{\partial J}{\partial \alpha_1}, \ldots, \frac{\partial J}{\partial \alpha_N} \right]_{\mathbf{f}^{(k)}}, \qquad [10]$$

where the object function is defined as $\mathbf{f} = \exp[j\boldsymbol{\varphi} - \boldsymbol{\alpha}]$, with $\boldsymbol{\varphi}$ representing the phase delay and $\boldsymbol{\alpha}$ the absorption. In what follows we denote the gradient by $\nabla$ for notational simplicity. We take the gradient of Eq. **9** and, by linearity of the derivation operation and denoting $H_i(\mathbf{f})$ by $\mathbf{H}_i = (H_1, \ldots, H_M)^T$, we get

$$\nabla J = \frac{1}{2} \sum_{i=1}^{N_v} \left[ \nabla(\mathbf{H}_i^T \mathbf{H}_i) - 2\nabla(\mathbf{H}_i^T \mathbf{g}_i) \right]. \qquad [11]$$

The term $\nabla(\mathbf{g}_i^T \mathbf{g}_i)$ is absent because measurements $\mathbf{g}_i$ do not depend on the estimate $\mathbf{f}$. Then, by the definition

$$\nabla \mathbf{H} = \begin{bmatrix} \frac{\partial H_1}{\partial \varphi_1} & \cdots & \frac{\partial H_1}{\partial \varphi_N} & \frac{\partial H_1}{\partial \alpha_1} & \cdots & \frac{\partial H_1}{\partial \alpha_N} \\ \cdots & & & & & \\ \frac{\partial H_M}{\partial \varphi_1} & \cdots & \frac{\partial H_M}{\partial \varphi_N} & \frac{\partial H_M}{\partial \alpha_1} & \cdots & \frac{\partial H_M}{\partial \alpha_N} \end{bmatrix}, \qquad [12]$$

we get

$$\nabla J = \sum_{i=1}^{N_v} \left[ \mathbf{H}_i^T \nabla \mathbf{H}_i - \mathbf{g}_i^T \nabla \mathbf{H}_i \right] \qquad [13]$$

$$= \sum_{i=1}^{N_v} \left[ \mathbf{r}_i^T \nabla \mathbf{H}_i \right], \qquad [14]$$

where $\mathbf{r}_i$ is the residual defined as $\mathbf{r}_i = \mathbf{H}_i - \mathbf{g}_i$. Finally, we get the expression required in Eq. **6**:

$$(\nabla J)^T = \sum_{i=1}^{N_v} (\nabla \mathbf{H}_i)^T \mathbf{r}_i. \quad [15]$$

In Eq. **15**, $(\nabla \mathbf{H}_i)^T$ is a matrix of size $(2N \times M)$ that is too large to be computed directly. Instead, we use a routine, described below, to calculate the vector $(\nabla \mathbf{H}_i)^T \mathbf{r}_i$ directly. We remind the reader of Eqs. **4** and **5** that describe the forward model where we drop the index $i$ to simplify the notation as the expression assumes the same form for each sample orientation:

$$\mathbf{H} = |\mathbf{u}_{\text{det}}|^2 \quad [16]$$

$$\mathbf{u}_{\text{det}} = F_d \text{diag}[\mathbf{f}_L] \dots F_{\Delta z} \text{diag}[\mathbf{f}_2] F_{\Delta z} \text{diag}[\mathbf{f}_1] \mathbf{u}_{\text{inc}}. \quad [17]$$

This forward operator allows for a convenient computation of the gradient by using a backpropagation scheme. We first calculate the gradient of Eq. **16**,

$$\nabla \mathbf{H} = \nabla |\mathbf{u}_{\text{det}}|^2 \quad [18]$$

$$= \nabla (\text{diag}[\mathbf{u}_{\text{det}}^*] \mathbf{u}_{\text{det}}) \quad [19]$$

$$= \text{diag}[\mathbf{u}_{\text{det}}] \nabla \mathbf{u}_{\text{det}}^* + \text{diag}[\mathbf{u}_{\text{det}}^*] \nabla \mathbf{u}_{\text{det}} \quad [20]$$

$$= 2\Re \{\text{diag}[\mathbf{u}_{\text{det}}^*] \nabla \mathbf{u}_{\text{det}}\}, \quad [21]$$

where the asterisk represents the complex conjugate. Thus,

$$(\nabla \mathbf{H})^\dagger \mathbf{r} = 2\Re \left\{ (\nabla \mathbf{u}_{\text{det}})^\dagger \mathbf{r}' \right\}, \quad [22]$$

where the dagger represents the Hermitian transpose and we have defined $\mathbf{r}' = \text{diag}[\mathbf{u}_{\text{det}}] \mathbf{r}$. Because it is not practical to compute the matrix $(\nabla \mathbf{u}_{\text{det}})^\dagger$ due to its size, we use a recursive scheme to compute $(\nabla \mathbf{u}_{\text{det}})^\dagger \mathbf{r}'$ directly. For that, we rewrite Eq. **17** as a recursive relationship for the optical field $\mathbf{u}_l$ just after layer $l$:

$$\mathbf{u}_1 = \text{diag}[\mathbf{f}_1] \mathbf{u}_{\text{inc}} \quad [23]$$

$$\mathbf{u}_l = \text{diag}[\mathbf{f}_l] F_{\Delta z} \mathbf{u}_{l-1} \quad [24]$$

$$\mathbf{u}_{\text{det}} = F_d \mathbf{u}_L. \quad [25]$$

The optical field $\mathbf{u}$ is thus known everywhere for a given object function $\mathbf{f}$. We then propagate the residual $\mathbf{r}'$ backward through the sample by using the same propagation relationships:

$$\mathbf{r}'_L = F_d^\dagger \mathbf{r}' \quad [26]$$

$$\mathbf{r}'_{l-1} = F_{\Delta z}^\dagger \text{diag}[\mathbf{f}_l]^\dagger \mathbf{r}'_l. \quad [27]$$

Note that the Fresnel operator is unitary; i.e., $F^\dagger = F^{-1}$. We take the gradient of Eqs. **23–25**:

$$\nabla \mathbf{u}_1 = \text{diag}[\mathbf{u}_{\text{inc}}] \nabla \mathbf{f}_1 \quad [28]$$

$$\nabla \mathbf{u}_l = \text{diag}[F_{\Delta z} \mathbf{u}_{l-1}] \nabla \mathbf{f}_l + \text{diag}[\mathbf{f}_l] F_{\Delta z} \nabla \mathbf{u}_{l-1} \quad [29]$$

$$\nabla \mathbf{u}_{\text{det}} = F_d \nabla \mathbf{u}_L. \quad [30]$$

We then take the Hermitian transpose and multiply by the residual; we get

$$(\nabla \mathbf{u}_1)^\dagger \mathbf{r}'_1 = (\nabla \mathbf{f}_1)^\dagger \text{diag}[\mathbf{u}_{\text{inc}}]^\dagger \mathbf{r}'_1 \quad [31]$$

$$(\nabla \mathbf{u}_l)^\dagger \mathbf{r}'_l = (\nabla \mathbf{f}_l)^\dagger \text{diag}[F_{\Delta z} \mathbf{u}_{l-1}]^\dagger \mathbf{r}'_l + + (\nabla \mathbf{u}_{l-1})^\dagger F_{\Delta z}^\dagger \text{diag}[\mathbf{f}_l]^\dagger \mathbf{r}'_l \quad [32]$$

$$(\nabla \mathbf{u}_{\text{det}})^\dagger \mathbf{r}' = (\nabla \mathbf{u}_L)^\dagger F_d^\dagger \mathbf{r}'. \quad [33]$$

We simplify the equations above by making use of Eqs. **24** and **25**,

$$(\nabla \mathbf{u}_1)^\dagger \mathbf{r}'_1 = (\nabla \mathbf{f}_1)^\dagger \text{diag}[\mathbf{u}_{\text{inc}}]^\dagger \mathbf{r}'_1 \quad [34]$$

$$(\nabla \mathbf{u}_l)^\dagger \mathbf{r}'_l = (\nabla \mathbf{u}_{l-1})^\dagger \mathbf{r}'_{l-1} + (\nabla \mathbf{f}_l)^\dagger \text{diag}[F_{\Delta z} \mathbf{u}_{l-1}]^\dagger \mathbf{r}'_l \quad [35]$$

$$(\nabla \mathbf{u}_{\text{det}})^\dagger \mathbf{r}' = (\nabla \mathbf{u}_L)^\dagger \mathbf{r}'_L, \quad [36]$$

which gives us a recursive relationship to calculate the gradient of the field at each layer. Note that $(\nabla \mathbf{f}_l)^\dagger$ is a matrix of size $2N \times M$ whose entries are nonzeros only for the diagonal entries corresponding to layer $l$ because $f_i$ depends only on $\alpha_i$ and $\varphi_i$. In practice, $(\nabla \mathbf{u}_{\text{det}})^\dagger \mathbf{r}'$ can be built layer by layer by stacking the second term of the right-hand side of Eq. **35** which reads, for pure phase objects $(\boldsymbol{\alpha} = \mathbf{0})$,

$$(\nabla \mathbf{f}_l)^\dagger \text{diag}[F_{\Delta z} \mathbf{u}_{l-1}]^\dagger \mathbf{r}'_l = -j \text{diag}[e^{-j\boldsymbol{\varphi}_l}] \text{diag}[F_{\Delta z} \mathbf{u}_{l-1}]^\dagger \mathbf{r}'_l \quad [37]$$

$$= -j \text{diag}[e^{-j\boldsymbol{\varphi}_l}] \text{diag}[e^{j\boldsymbol{\varphi}_l}] \text{diag}[\mathbf{u}_l^*] \mathbf{r}'_l \quad [38]$$

$$= -j \text{diag}[\mathbf{u}_l^*] \mathbf{r}'_l, \quad [39]$$

where we have used Eq. **24**. Finally, according to Eq. **22**, we obtain layer $l$ of $(\nabla \mathbf{H})^\dagger \mathbf{r}$,

$$(\nabla \mathbf{H})^\dagger \mathbf{r}|_{\text{layer } l} = 2\Im \left\{ \text{diag}[\mathbf{u}_l^*] \mathbf{r}'_l \right\}, \quad [40]$$

where $\Im$ denotes the imaginary part.

## Materials and Methods

The experimental apparatus is shown in Fig. 2*A*. The light source is a CW He-Ne laser at 632.8 nm that is spatially filtered, expanded, and collimated into a quasi-plane wave with an Airy disk intensity profile of 33 mm in diameter. The sample is mounted on a 2-axis rotation stage rotating along the *x* and *y* axes. The sample middle plane is imaged using a demagnifying telescope (×0.50) lens system to enhance the effect of diffraction and increase the field of view on the detector. The detector is an EM-CCD (QImaging Rolera EM-C2) with a 1,004 × 1,002 array of 8 × 8-μm pixels. To simulate the diffraction occurring in an X-ray measurement, the detector is defocused by a distance $\Delta z = 58$ mm from the image plane, which corresponds to Fresnel numbers ranging from 0.7 to 5.5 for the different object features.

The sample corresponds to a $10^4\times$ scale-up of a real IC design. The original IC comprises 13 layers, including the doped layers. We kept only layers 5 to 8 from the original design (relabeled here 1 to 4) shown in Fig. 2*C* that contain copper patterns that would induce a significant phase delay in the X-ray regime. The 4 glass plates corresponding to the IC layers were cut in a 500-μm-thick fused silica wafer and $575 \pm 5$-nm deep patterns (measured after fabrication with a Bruker DekTak XT stylus profilometer) were obtained by wet etching. To control the phase contrast and reduce parasitic reflections between the layers, we used an immersion oil (Fig. 2*B*) with a refractive index $n_D = 1.400 \pm 0.0002$ at 25 °C from Cargille-Sacher Laboratories. According to the manufacturer, the refractive index of the oil is $n_{\text{oil}} = 1.4005 \pm 0.0002$ at 632.8 nm and 20 °C. The refractive index of fused silica is $n_{\text{glass}} = 1.457$ at 632.8 nm and 20 °C (65), which gives a contrast of $\Delta n = 0.0565 \pm 0.0005$. The corresponding phase shift for a single pattern is then $\Delta\varphi = kd\Delta n = 0.323 \pm 0.006$ rad.

The sample layers are fabricated on double-sided polished 150-mm-diameter and 500-μm-thick fused silica wafers. A 1-μm-thick positive tone resist (Megaposit SPR700) is spin coated at 3,500 rpm on both sides of the wafer and soft baked at 95 °C for 30 min in a convection oven. The backside was also coated for protection from the forthcoming wet etch. Scaled versions of IC designs in GDSII format are then patterned directly using a maskless aligner (MLA150; Heidelberg Instruments) with a 405-nm laser and developed using an alkaline developer (Shipley Microposit MF CD-26) for 45 s followed by a deionized (DI) water rinse and $N_2$ drying. A hard bake at 120 °C for 30 min is carried out to stabilize the patterned features. A short descum of 2 min at 1,000-W and 0.1-Torr $O_2$ pressure in a barrel asher is also performed to remove any resist residue. The wafers are subsequently etched for 7 min at a rate of ∼80 nm/min in buffered oxide etch. The resist is then stripped from the wafer by a long ash (1 h) followed by a Piranha

clean (3:1 $H_2SO_4$:$H_2O_2$), DI water rinse, and $N_2$ drying. Finally, the wafers are diced into squares of 50 mm $\times$ 50 mm and cleaned again with Piranha, DI water rinse, and $N_2$ drying.

The computation of the approximant is performed with the MATLAB software on an Intel i9-7900X processor running at 3.3 GHz. The DNN training and testing are performed under Keras with Tensorflow backend running on an NVIDIA Titan Xp graphics processing unit.

**ACKNOWLEDGMENTS.** This work was supported by the Intelligence Advanced Research Projects Activity (FA8650-17-C-9113).

1. R. N. Bracewell, Aerial smoothing in radio astronomy. *Aust. J. Phys.* **7**, 615–640 (1954).
2. R. N. Bracewell, Strip integration in radio astronomy. *Aust. J. Phys.* **9**, 198–217 (1956).
3. R. N. Bracewell, Two-dimensional aerial smoothing in radio astronomy. *Aust. J. Phys.* **9**, 297–314 (1956).
4. A. M. Cormack, Representation of a function by its line integrals, with some radiological applications. *J. Appl. Phys.* **34**, 2722–2727 (1963).
5. A. M. Cormack, Representation of a function by its line integrals, with some radiological applications. II. *J. Appl. Phys.* **35**, 2908–2913 (1964).
6. G. N. Hounsfield, Computerized transverse axial scanning (tomography): Part I. Description of system. *Br. J. Radiol.* **46**, 1016–1022 (1973).
7. J. Ambrose, Computerized transverse axial scanning (tomography): Part II. Clinical application. *Br. J. Radiol.* **46**, 1023–1047 (1973).
8. D. J. de Rosier, A. Klug, The reconstruction of a three dimensional structure from projections and its application to electron microscopy. *Proc. R. Soc. Lond. A* **317**, 130–134 (1968).
9. D. J. de Rosier, A. Klug, Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 319–340 (1970).
10. R. A. Crowther, L. A. Amos, J. T. Finch, D. J. de Rosier, A. Klug, Three dimensional reconstructions of spherical viruses by Fourier synthesis from electron micrographs. *Nature* **226**, 421–425 (1970).
11. J. Radon, Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten. *Ber. Sachsische Acad. Wiss.* **69**, 262–267 (1917).
12. B. K. P. Horn, Density reconstruction using arbitrary ray-sampling schemes. *Proc. IEEE* **66**, 551–562 (1978).
13. B. K. P. Horn, Fan-beam reconstruction methods. *Proc. IEEE* **67**, 1616–1623 (1979).
14. J. T. Dobbins III, D. J. Godfrey, Digital x-ray tomosynthesis: Current state of the art and clinical potential. *Phys. Med. Biol.* **48**, R65–R106 (2003).
15. D. W. Sweeney, C. M. Vest, Reconstruction of three-dimensional refractive index fields by holographic interferometry. *Appl. Opt.* **11**, 205–207 (1972).
16. C. M. Vest, *Holographic Interferometry* (Wiley, 1979).
17. J. C. Petruccelli, L. Tian, G. Barbastathis, The transport of intensity equation for optical path length recovery using partially coherent illumination. *Opt. Exp.* **21**, 14430 (2013).
18. E. Wolf, Three-dimensional structure determination of semi-transparent objects from holographic data. *Opt. Commun.* **1**, 153–156 (1969).
19. W. Choi *et al.*, Tomographic phase microscopy. *Nat. Methods* **4**, 717–719 (2007).
20. A. N. Tikhonov, On the solution of ill-posed problems and the method of regularization. *Dokl. Acad. Nauk SSSR* **151**, 501–504 (1963).
21. A. N. Tikhonov, On the stability of algorithms for the solution of degenerate systems of linear algebraic equations. *Zh. Vychisl. Mat. i Mat. Fiz.* **5**, 718–722 (1965).
22. E. J. Candès, T. Tao, Decoding by linear programming. *IEEE Trans. Inf. Theory* **51**, 4203–4215 (2005).
23. E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete Fourier information. *IEEE Trans. Inf. Theory* **52**, 489–509 (2006).
24. D. L. Donoho, Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
25. E. J. Candès, J. Romberg, T. Tao, Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–1223 (2006).
26. E. J. Candès, T. Tao, Near optimal signal recovery from random projections: Universal encoding strategies?*IEEE Trans. Inf. Theory* **52**, 5406–5425 (2006).
27. D. J. Brady, A. Mrozack, K. MacCabe, P. Llull, Compressive tomography. *Adv. Opt. Photon* **7**, 756–813 (2015).
28. M. Elad, M. Aharon, "Image denoising via learned dictionaries and sparse representation" in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (IEEE Computer Society, Washington, DC, 2006), vol. 1, pp. 895–900.
29. M. Aharon, M. Elad, A. Bruckstein, K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.* **54**, 4311–4322 (2006).
30. Y. LeCun, Y. Bengio, G. Hinton, Deep learning. *Nature* **521**, 436–444 (2015).
31. K. Gregor, Y. LeCun, "Learning fast approximations of sparse coding" in *Proceedings of the 27th International Conference on International Conference on Machine Learning, ICML'10*, J. Fürnkranz, T. Joachims, Eds. (International Conference on Machine Learning, La Jolla, CA, 2010), pp. 399–406.
32. M. Mardani *et al.*, Recurrent generative adversarial networks for proximal learning and automated compressive image recovery. arXiv:1711.10046 (27 November 2017).
33. A. Sinha, J. Lee, S. Li, G. Barbastathis, Lensless computational imaging through deep learning. *Optica* **4**, 1117–1125 (2017).
34. Y. Rivenson, Y. Zhang, H. Günaydin, D. Teng, A. Ozcan, Phase recovery and holographic image reconstruction using deep learning in neural networks. *Light Sci. Appl.* **7**, 17141 (2018).
35. S. Li, 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).
36. Z. D. C. Kemp, Propagation based phase retrieval of simulated intensity measurements using artificial neural networks. *J. Opt.* **20**, 045606 (2018).
37. T. Shimobaba, T. Kakue, T. Ito, Convolutional neural network-based regression for depth prediction in digital holography. arXiv:1802.00664 (2 February 2018).
38. Z. Ren, Z. Xu, E. Y. Lam, Autofocusing in digital holography using deep learning. *Proc. SPIE* **10499**, 104991V (2018).
39. Y. Wu *et al.*, Extended depth-of-field in holographic image reconstruction using deep learning based auto-focusing and phase-recovery. *Optica* **5**, 704–710 (2018).
40. C. Dong, C. Loy, K. He, X. Tang, "Learning a deep convolutional neural network for image super-resolution" in *European Conference on Computer Vision (ECCV), Part IV/Lecture Notes on Computer Science*, D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars, Eds. (Springer International Publishing, Cham, Switzerland, 2014), vol. 8692, pp. 184–199.
41. C. Dong, C. Loy, K. He, X. Tang, Image super-resolution using deep convolutional networks. *IEEE Trans. Patt. Anal. Mach. Intell.* **38**, 295–307 (2015).
42. J. Johnson, A. Alahi, L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution" in *European Conference on Computer Vision (ECCV)/Lecture Notes on Computer Science*, B. Leide, J. Matas, N. Sebe, M. Welling, Eds. (Springer International Publishing, Cham, Switzerland, 2016), vol. 9906, pp. 694–711.
43. Y. Rivenson *et al.*, Deep learning microscopy. *Optica* **4**, 1437–1443 (2017).
44. M. Lyu, H. Wang, G. Li, G. Situ, Exploit imaging through opaque wall via deep learning. arXiv:1708.07881 (9 August 2017).
45. S. Li, M. Deng, J. Lee, A. Sinha, G. Barbastathis, Imaging through glass diffusers using densely connected convolutional networks. *Optica* **5**, 803–813 (2018).
46. N. Borhani, E. Kakkava, C. Moser, D. Psaltis, Learning to see through multimode fibers. *Optica* **5**, 960–966 (2018).
47. Y. Li, Y. Xue, L. Tian, Deep speckle correlation: A deep learning approach toward scalable imaging through scattering media. *Optica* **5**, 1181–1190 (2018).
48. M. Lyu *et al.*, Deep-learning-based ghost imaging. *Sci. Rep.* **7**, 17865 (2017).
49. C. Chen, Q. Chen, J. Xu, V. Koltun, "Learning to see in the dark" in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, New York, NY, 2018), pp. 3291–3300.
50. A. Goy, K. Arthur, S. Li, G. Barbastathis, Low photon count phase retrieval using deep learning. arXiv:1806.10029 (25 June 2018).
51. M. Mardani *et al.*, Deep generative adversarial networks for compressed sensing automates MRI. arXiv:1706.00051 (31 May 2018).
52. J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, D. Rueckert, A deep cascade of convolutional neural networks for MR image reconstruction. arXiv:1703.00555 (1 March 2017).
53. K. Hwan Jin, M. T. McCann, E. Froustey, M. Unser, Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Image Proc.* **26**, 4509–4521 (2017).
54. M. T. McCann, K. Hwan Jin, M. Unser, Convolutional neural networks for inverse problems in imaging: A review. *IEEE Sig. Process. Mag.* **34**, 85–95 (2017).
55. H. Gupta, K. Hwan Jin, H. Q. Nguyen, M. T. McCann, M. Unser, CNN-based projected gradient descent for consistent CT image reconstruction. *IEEE Trans. Med. Imaging* **37**, 1440–1453 (2018).
56. Y. Sun, Z. Xia, U. S. Kamilov, Efficient and accurate inversion of multiple scattering with deep learning. *Opt. Exp.* **26**, 14678–14688 (2018).
57. T. Nguyen, V. Bui, G. Nehmetallah, Computational optical tomography using 3-d deep convolutional neural networks. *Opt. Eng.* **57**, 043111 (2017).
58. M. D. Feit, J. A. Fleck Jr, Light propagation in graded-index optical fibers. *Appl. Opt.* **17**, 8–12 (1978).
59. M. A. Mak, Trusted defense microelectronics: Future access and capabilities are uncertain. https://www.gao.gov/assets/680/673401.pdf. Accessed 6 September 2019.
60. U. S. Kamilov *et al.*, Learning approach to optical tomography. *Optica* **2**, 517–522 (2015).
61. U. S. Kamilov *et al.*, Optical tomographic image reconstruction based on beam propagation and sparse regularization. *IEEE Trans. Comput. Imaging* **2**, 59–70 (2016).
62. G. Huang, Z. Liu, K. Q. Weinberger, Densely connected convolutional networks. arXiv:1608.06993 (25 August 2016).
63. A. Beck, M. Teboulle, A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2**, 183–202 (2009).
64. A. Beck, M. Teboulle, Fast gradient-based algorithm for constrained total variation image denoising and deblurring problems. *IEEE Trans. Image Process.* **18**, 2419–2434 (2009).
65. I. H. Malitson, Interspecimen comparison of the refractive index of fused silica. *J. Opt. Soc. Am.* **55**, 1205–1209 (1965).