

# Advantage of Machine Learning over Maximum Likelihood in Limited-Angle Low-Photon X-Ray Tomography

Zhen Guo, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA; Jung Ki Song, Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02139, USA; George Barbastathis, Singapore-MIT Alliance for Research and Technology (SMART) Centre, Singapore 138002; Michael E. Glinsky, Courtenay T. Vaughan, Kurt W. Larson, Sandia National Laboratory Albuquerque, New Mexico, 87123, USA; Bradley K. Alpert, Applied and Computational Mathematics Division, National Institute of Standards and Technology, Boulder, Colorado, 80305, USA; Zachary H. Levine, Quantum Measurement Division, National Institute of Standards and Technology, Gaithersburg, Maryland 20899, USA

## Abstract

Limited-angle X-ray tomography reconstruction is an ill-conditioned inverse problem in general. Especially when the projection angles are limited and the measurements are taken in a photon-limited condition, reconstructions from classical algorithms such as filtered backprojection may lose fidelity and acquire artifacts due to the missing-cone problem. To obtain satisfactory reconstruction results, prior assumptions, such as total variation minimization and nonlocal image similarity, are usually incorporated within the reconstruction algorithm. In this work, we introduce deep neural networks to determine and apply a prior distribution in the reconstruction process. Our neural networks learn the prior directly from synthetic training samples. The neural nets thus obtain a prior distribution that is specific to the class of objects we are interested in reconstructing. In particular, we used deep generative models with 3D convolutional layers and 3D attention layers which are trained on 3D synthetic integrated circuit (IC) data from a model dubbed CircuitFaker. We demonstrate that, when the projection angles and photon budgets are limited, the priors from our deep generative models can dramatically improve the IC reconstruction quality on synthetic data compared with maximum likelihood estimation. Training the deep generative models with synthetic IC data from CircuitFaker illustrates the capabilities of the learned prior from machine learning. We expect that if the process were reproduced with experimental data, the advantage of the machine learning would persist. The advantages of machine learning in limited angle X-ray tomography may further enable applications in low-photon nanoscale imaging.

## Background

Limited-angle X-ray tomography has the ability to image the interior of three-dimensional (3D) objects non-invasively without collecting the measurements from a full set of rotation angles. It has drawn wide attention in practical nanoscale imaging due to its advantage in having a relatively short data acquisition time. Also, in some cases, not all angles are physically accessible.

Typically, the tomographic imaging system consists of a sample holder, an objective zone plate, and a CCD camera, and the illumination is generated from the X-ray source. The measurements are taken from a series of rotational angles with respect to the sample of interest, where a cone-beam geometry is generally assumed to produce the ray projection from the source point to the sample, and then from sample to the center of each

detector pixel. After the measurements, objects subsequently can be reconstructed based on 3D computed tomography algorithms. Theoretically, full-angle measurement is preferred to avoid the missing cone problem in the reconstruction process. In practice, however, limited-angle measurement is often used due to the time of acquiring the full angle measurement. For objects or samples that are radiation sensitive, a low-photon budget per scan is also preferred to minimize the total exposure and potential damage.

Limited-angle X-ray tomography is an ill-conditioned inverse problem [1, 2], where the goal is to find a discrete representation of an object  $f$  based on the observations  $g$  taken on a digital camera at limited number of angles. When limited-angle tomography is applied where the illumination is limited to only a few photons, the noise sensitivity of poor conditioning becomes even more evident: not only do the collected Fourier slices cover the Fourier space unevenly due to the limited scans, they are also inaccurate due to the presence of shot noise in the measurements. Limited angle X-ray tomography therefore has had limited utility for radiation sensitive samples. To solve the limited angle X-ray tomography with photon scarcity, regularization is required during the reconstruction process. Some improvement can be obtained by extrapolating missing data [3]. Data consistency conditions, e.g., the Helgason-Ludwig consistency conditions, further improves the quality of extrapolation [4]. Still, extrapolating methods are struggling with complex structures and are not robust to noise in the experimental measurements.

Iterative algorithms with constraints known a priori reconstruct the object using multiple steps. The algorithms start with an assumed object, simulate the measurement from the assumed object, compare the experimental measurements and simulated measurements, and then update the object based on the difference between measurements and simulations. The last step also includes discrepancy in the prior terms into the computation of the update. The process continues until a certain convergence criterion is achieved. Iterative algorithms with prior constraints such as total variation minimization and nonlocal image similarity often exhibit improved resilience to noise. Previously, we also have shown a Bayesian approach [5] based on the Bouman-Sauer formulation for the iterative reconstruction algorithm [6]. The regularization constrains the iterative optimization to a subdomain in which the object belongs, thereby effectively improving the reconstruction quality [7, 8].

Recently, machine learning has been applied successfully to

limited-angle tomography to overcome the challenge in solving the inverse problem. Efforts have been made in using learned priors to provide information of missing data during reconstruction [9, 10, 11]. In particular, deep learning, a subset of machine learning that is based on artificial neural networks with representation learning, achieved promising results [11, 12, 13].

For tomographic inverse problems, U-net-like neural network architectures [14] have been widely used to produce reconstruction pixel-by-pixel. For example, U-net has been used to predict invisible singularities in the image object [12], to generate missing projections with a data-consistent reconstruction method [13], and to improve the reconstruction quality from the conventional FBP method [11]. Unlike the general priors in iterative algorithms, the learned prior from a deep learning method is conditioned on a large amount of paired training data. Therefore, the learned prior explores the statistical property of the training distribution.

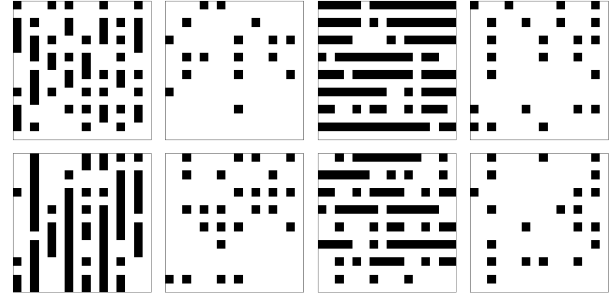
In this work, we extend the application of limited-angle X-ray tomography to experimental conditions of low-photon incidence. To approximate the resulting ill-conditioned inverse problem and to obtain a satisfactory reconstruction quality, we apply machine learning to determine a prior distribution for the reconstruction process. In particular, we use deep generative models with 3D convolution and 3D attention which are trained on 3D synthetic integrated circuit (IC) data from a model dubbed CircuitFaker. We demonstrate that the priors from our deep generative models drastically improve the IC reconstruction quality on synthetic data from maximum likelihood estimation when the projection angles and photon budgets are limited. Beyond existing research that uses machine learning for limited angle X-ray tomography, our generative model exhibits improvements over maximum likelihood reconstructions under low-photon conditions. We further examine different neural network architectures to reveal some of the important network design choices for solving the inverse problem.

## Method

The overall pipeline of our method is organized as follows: First, we generate synthetic integrated circuit (IC) layouts using CircuitFaker, an in-house model. Then, we simulate the limited-angle X-ray tomography radiographs. Next, we apply the maximum likelihood method to generate an initial IC reconstruction. Finally, we feed the initial IC reconstruction to the (trained) deep generative model and compare the reconstruction quality of the test set by calculating the bit error rate.

### CircuitFaker

CircuitFaker is an algorithm that generates a random set of voxels with binary values which resembles an integrated circuit interconnect. The synthetic circuits from CircuitFaker is the class of artificial objects for tomographic reconstruction, and the implicit correlations in their spatial features are the priors to be assumed or to be learned for the inverse algorithms. A particular draw of CircuitFaker assigns a bit in each of  $N = N_x N_y N_z$  locations. These locations are indexed as  $i_\ell = 1, \dots, N_\ell$ , with  $\ell = 1, 2, 3$  for  $x, y$ , and  $z$ . All bits are initialized to 0. In the first round, there are wire seed points for all locations  $(i_1, i_2, i_3)$  with  $i_1, \dots, i_3$  odd. Each seed point is set by a Bernoulli draw with probability  $p_w$  of getting a 1. There are three kinds of layers,  $x, y$ , and via



**Figure 1.** Selected  $16 \times 16 \times 8$  circuit from CircuitFaker. Each image is a slice of 2D layer in the  $z$  dimension. The value of  $z$  increases as a raster scan of the 8 slices shown. Black indicates copper and white indicates silicon. Here,  $x$  layers are the first and fifth layers in  $z$ ,  $y$  layers are the third and seventh layers in  $z$ . Others are via layers.

layers. The  $x$  wiring layers have index  $i_3 = 1 \bmod 4$ . The  $y$  wiring layers have  $i_3 = 3 \bmod 4$ . The via layers are the others, i.e.,  $i_3$  even. In the second round, a point on an  $x$  wiring layer to the immediate right of a point with value 1 is set to 1 with probability  $p_x$ . A point on a  $y$  wiring layer immediately above in plan view a point with a value 1 is set to 1 with probability  $p_y$ . Similarly, a point on a via layer immediately above a point with a value 1 is set to 1 with probability  $p_z$ . In this paper, we chose these parameters:  $N_x = N_y = 16$ ,  $N_z = 8$ ,  $p_w = 0.75$ ,  $p_x = p_y = 0.8$ , and  $p_z = 0.5$ . Fig. 1 shows one of the generated circuits with size  $16 \times 16 \times 8$ .

### Imaging geometry for X-ray tomography

We assumed that each voxel in the circuit is in size of size  $0.15 \mu\text{m} \times 0.15 \mu\text{m} \times 0.30 \mu\text{m}$ . Therefore, the total volume of the circuit is  $2.4 \mu\text{m} \times 2.4 \mu\text{m} \times 2.4 \mu\text{m}$ . The detector is assumed to be in the  $x$ - $z$  plane at a tilt angle of  $\varphi = 0^\circ$ . The rotation axis is  $z$ . The detector is  $13.44 \text{ mm} \times 13.44 \text{ mm}$  with  $32 \times 32$  pixels of size  $420 \mu\text{m} \times 420 \mu\text{m}$ . The system operates with a geometric magnification of 5000, with a source-sample distance of  $10 \mu\text{m}$ . There are 8 tilt angles from  $-30^\circ$  to  $+22.5^\circ$  with an increment of  $7.5^\circ$ . There is a single source point with a cone-beam geometry. A single ray is taken from the source point to the center of each detector pixel. Small corrections for variations in the source-to-detector pixel distance, the obliquity, and the source's Heel effect are neglected. The exact detection is written as:

$$g^{(0)} = \int dE D(E) I^{(0)}(E) e^{-\alpha(E)A f}. \quad (1)$$

Here,  $A$  is the system matrix, i.e., the distance each projection makes from the source to a detector pixel for all the angles,  $f$  is the vector of the object,  $E$  is the photon energy,  $\alpha(E)$  is the absorption coefficient at energy  $E$  for copper,  $I^{(0)}(E)$  is the source intensity,  $D(E)$  is the detector efficiency.  $g^{(0)}$  is the vector of expected measurements on the camera in number of photons for all the angles. With the consideration of Poisson statistics, the final measurements  $g$  is:

$$g \sim \mathcal{P}(g^{(0)}) \quad (2)$$

where  $\mathcal{P}(\lambda)$  denotes Poisson sampling with parameter  $\lambda$ . In this work, we do not use scatter corrections and we are restricted to a single material, namely copper, at its bulk density of  $8.960 \text{ g/cm}^3$ .

The spectrum consists of two equally weighted lines at 9362 eV and 9442 eV, the Pt  $L_{\alpha}$  fluorescence lines. The attenuation per voxel is about 2 % if copper is present. The exact value depends on the details of how a ray intersects a voxel. The specific example was chosen to support an experimental project to perform integrated circuit tomography with a laboratory scale instrument [15].

### Maximum likelihood estimation

Maximum likelihood estimation as applied to projective tomography is well-known. The method produces the maximum log-likelihood reconstruction circuit  $\tilde{f}$  with a given set of tomographic measurements  $g$  (in the number of photon counts for each detector pixel) at different angles.  $f$  is the voxelized binary matrix representing the 3D circuit. A voxel value of 1 indicates the presence of copper and 0 indicates silicon. The objective is to find an optimal  $\tilde{f}$  given the measurements  $g$ :

$$\tilde{f}(g) = \arg \max_{f^{(0)}} \left[ L_{\text{ML}}(g|f^{(0)}) + \Psi(f^{(0)}) \right] \quad (3)$$

$$L_{\text{ML}}(g|f^{(0)}) = \sum_i \left[ \ln g_i! - g_i \ln g_i^{(0)} + g_i^{(0)} \right]. \quad (4)$$

Here  $L_{\text{ML}}$  is the log-likelihood under the assumption of Poisson statistics,  $\Psi$  is a regularization function or log of the Bayesian prior,  $g^{(0)}$  is the simulated measurement from a proposed object  $f^{(0)}$ ,  $\sum_i$  sums over all individual measurements, and  $\tilde{f}$  is the optimal reconstruction based on maximizing the log likelihood [5, 6].

### Bit error rate calculation

The bit error rate (BER) is introduced as an evaluation metric to assess the performance of the algorithms. It provides a measure of the frequency of misclassification for binary values in the voxels in a given circuit. That is, BER is the probability of classifying a specific voxel in a circuit to be 1 while the ground truth value for the corresponding voxel is actually 0 and vice versa. The procedure for computing bit error rate in this paper is slightly modified from the standard used in communication theory, and is as follows:

1. compute posterior distributions  $p(f=0|\tilde{f})$  by multiplying the probability density functions (PDFs)  $p(\tilde{f}|f=0)$  and  $p(\tilde{f}|f=1)$  and their corresponding prior distributions ( $p_0$  and  $p_1$ ). Here,  $f=0$  is the set of silicon voxels in the circuit,  $f=1$  is the set of copper voxels in the circuit;
2. apply a chosen threshold to classify 0 and 1 based on likelihood functions;
3. compute the error rates for 0 and 1 ( $\eta_0$  and  $\eta_1$ , respectively) by summing over the misclassified region in the probability density functions;
4. derive an average bit error rate:  $\eta_{\text{avg}} = \eta_0 p_0 + \eta_1 p_1$ .

### Deep generative models

A deep generative model, known as a conditional generative adversarial network (cGAN) [16], is a supervised machine learning technique that we used to improve the 3D IC reconstruction over maximum likelihood estimation. When the projection angles and photons per ray are limited, the reconstructions from maximum likelihood estimation contain artifacts due to the missing cone problem and photon shot noise. The bit error rate of these reconstructions would eventually drop below the acceptable thresh-

old of single error per sample when the angular range and/or photon per ray of the tomographic measurements decrease. The deep generative model in our paper improves the Maximum Likelihood reconstructions in these situations through inference of what the likeliest true IC object structure is based on the (noisy) observation and, crucially, *what it has learned through its training*. We denote the noisy maximum likelihood estimate  $\tilde{f}$  as the Approximant.

In total, four variants of deep generative models have been investigated. Those variants are: deep generative model (baseline), generative model with axial attention, generative model with a scattering representation, and generative axial model with a scattering representation.

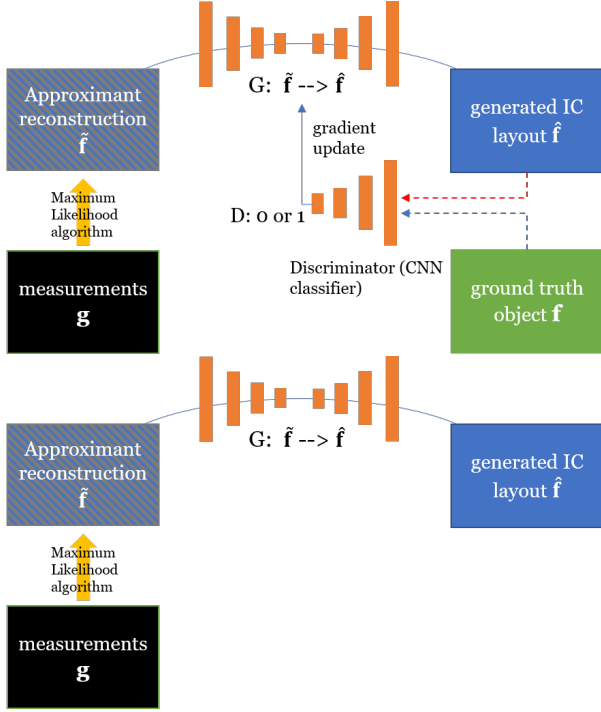
The first generative model applies the learned prior  $p(f)$  to the input distribution  $p(\tilde{f}|f)$ , and estimates  $p(f|\tilde{f})$  using Bayes' theorem [17, 18], where  $p(f)$  describes a distribution of noiseless IC samples. That is, the generative model learns the prior using a combination of generator and discriminator, which are trained to compete with each other until they reach the Nash equilibrium. We note the output of the GAN in equilibrium as  $\hat{f}$ . The entire generative model is formulated as follows [16]:

$$\arg \min_G \max_D \left( \mathbb{E}_f [\log D(f)] + \mathbb{E}_{\tilde{f}} [\log (1 - D(G(\tilde{f})))] \right) \quad (5)$$

The generator in our model is a 3D autoencoder that first learns to convert IC to latent space representation using 3D convolutional encoder, and then decodes the representation back to IC. The discriminator is a 3D convolutional classifier that tries to find whether the output from the generator is realistic or not. Both generator and discriminator in our model have convolutional kernels that are spectrally normalized to stabilize the training process [19]. This serves as the baseline generative model for our paper.

Our second variant of generative model is based on axial attention that harvests the contextual information in the input reconstruction. To build such model, we replace the 3D convolution in the encoder of the generator with full axial attention to extract or detect global features in the input reconstructions. This technique reduces the computational complexity of non-local attention to  $\mathcal{O}(hwzm)$ , where  $h$ ,  $w$ ,  $z$ , are the height, width, and depth of the input features, respectively,  $m$  is the local constraint constant [20]. Therefore, a network with axial attention is more efficient than a self-attention network, enabling long-range and global feature learning to overcome the limitation of locality in convolutional networks.

The third and fourth variants of generative models include the wavelet scattering transform [21, 22] of the IC reconstruction as additional input to the model. A scattering representation can be produced without training, and the produced representation is capable of including features at multiple scales. When combined with the renormalization technique, the generative model would be further conditioned on the scattering representation in generating realistic IC layout. In particular, the wavelet representation, after being fed to a trainable transformation of a fully connected layer, re-scales and re-shifts the normalized feature values from convolution. This way, the information about global features of the IC will be embedded into the reconstruction process of the neural network [23], which may help the network in learning the mapping from noisy reconstruction to noiseless reconstruction.



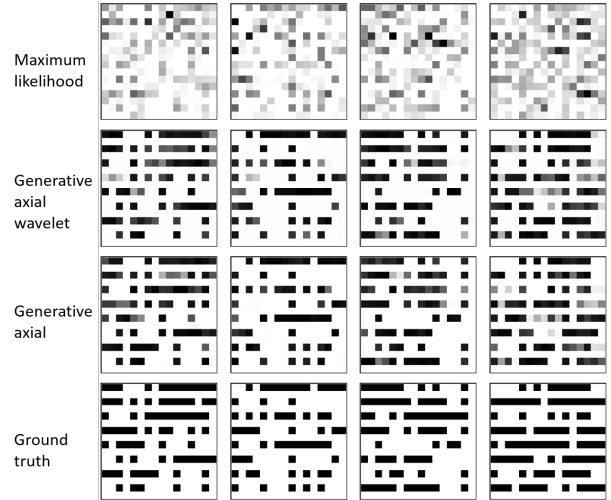
**Figure 2.** (a) Supervised training of the generative model with pairs of  $(f, g)$ . (b) Testing/operation on samples never used during training.

## Network training

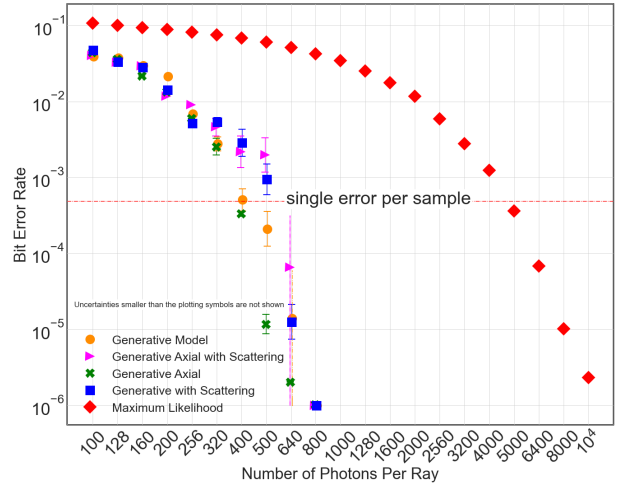
Our proposed networks were implemented in Python 3.7.9 using TensorFlow 2.3.1, and trained with an NVIDIA V100 tensor core graphics processing unit on MIT Supercloud [24]. An Adam optimizer [25] was used with parameters  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . Two time-scale update rule (TTUR) technique is used to stabilize the training of generative network [26, 27], where the initial learning rate was  $10^{-4}$  for the generator and  $4 \times 10^{-4}$  for discriminator. In each iteration, generator would be updated four times while discriminator would be updated once. Training sets of 1800 reconstructions were generated independently for each condition studied, except that the ground truth was common. An additional 200 reconstructions per condition were used for testing. The learning rate would be reduced by half when the validation loss stops improving for 5 epochs. We set the maximum number of iterations to be 200, and the training would stop early when either the validation loss plateaus for 20 epochs, or the minimal learning rate  $10^{-8}$  is reached. This early-stop technique would prevent the model from over-fitting. The loss function for the autoencoder/generator consists of two parts: supervised loss and adversarial loss. We choose supervised loss to be the negative of the Pearson correlation coefficient  $-r_{f, \tilde{f}}$ , which is defined as

$$-r_{f, \tilde{f}} = -\frac{\text{cov}(f, \tilde{f})}{\sigma_f \sigma_{\tilde{f}}}, \quad (6)$$

where cov is the covariance,  $\sigma$  is the standard deviation. The total objective of training is to find the optimal generator  $G_{\text{opt}}$  given



**Figure 3.** Selected Examples of IC Reconstructions for Different Methods



**Figure 4.** Maximum Likelihood vs. Generative Model Reconstructions with  $-30^\circ$  to  $22.5^\circ$  angular range.

the Approximant  $\tilde{f}$ :

$$G_{\text{opt}}(\tilde{f}) = \mathbb{E}_{f, \tilde{f}}[-r_{f, G(\tilde{f})}] + \lambda \arg \min_G \max_D \left( \mathbb{E}_f[\log D(f)] + \mathbb{E}_{\tilde{f}}[\log(1 - D(G(\tilde{f})))] \right) \quad (7)$$

The hyper-parameter  $\lambda$  that controls the degree of generation from input noise to features.

## Results

Fig. 3 shows selected examples of IC reconstructions at limited-angle and low photon conditions. Each row represents different reconstruction methods, and each column represents the same location at the given IC distribution. The last row represents the IC ground truth. Drastic improvement is visible when comparing the maximum likelihood reconstructions and generative reconstructions. The numerical result of limited angle and low photon tomography is shown in Fig. 4. The  $x$  axis is the number of photons per ray in the tomographic projection,  $y$  axis is the bit error rate of the reconstructed 3D IC samples. The angular

range is fixed at  $-30^\circ$  to  $22.5^\circ$  with  $7.5^\circ$  steps for all. There are five reconstructions we are comparing: reconstruction based on maximum likelihood estimation, and reconstruction based on four variants of generative model. The transition from above a single-error-per-sample to below happens between 320 to 640 photons per ray. Each of the simulations for the generative model is done with 1800 training sets and 200 test sets, and IC data is  $16 \times 16 \times 8$  voxels. For the transition cases between 320 and 640 photons per ray, we repeat the simulation with a total of five independent synthetic sets of IC circuits and report the means and standard errors in the plot. With 640 photons per ray, all the bit error rates from generative model reconstructions drop at least two orders of magnitude relative to the maximum likelihood reconstructions. In particular, the generative model with axial attention performs the best in terms of its lower mean and standard error, reaching a single error per sample at 400 photons per ray. We may attribute this to the application of axial attention to capture long range interactions within the input. The maximum likelihood reconstructions reach a single error per sample when the number of photons per ray is 5000. Therefore, in simulation, the generative model using machine learning can reduce the photon budget one order of magnitude to reach a single error per sample.

## Conclusion

In this work, we applied machine learning to limited-angle X-ray tomography with low-photon imaging conditions. Our deep-learning model is generative, and it is based on 3D convolutional layers and 3D attention layers, and then trained on 3D synthetic integrated circuit (IC) data from CircuitFaker. We demonstrated that the prior distributions from our deep generative models can dramatically improve the IC reconstruction quality on synthetic data compared to a maximum likelihood estimation algorithm when the projection angles and photon budgets are limited. Training the deep generative models with synthetic IC data from CircuitFaker illustrates the capabilities of the prior from machine learning. We expect that if the process was reproduced with experimental data, the advantage of the machine learning would persist. The advantages of using machine learning in limited angle X-ray tomography may further enable applications in low-photon nanoscale imaging.

## Acknowledgement

The authors thank Zachary Grey for extensive comments on an earlier draft. This research at MIT was supported by IARPA contract number FA8050-17-C-9113. G.B. also acknowledges support by Singapore's National Research Foundation through the Intra-Create grant NRF2019-THE002-0006 "Retinal Analytics via Machine learning aiding Physics (RAMP)." Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia LLC (NTESS), a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration (NNSA) under contract DE-NA0003525. This paper describes objective technical results and analysis. Any subjective views or opinions that might be expressed in the paper do not necessarily represent the views of the U.S. Department of Energy or the United States Government. (Sandia technical report number SAND2021-15957 C). Research at NIST of B.K.A. and Z.H.L. supported in part by the Office of the Director

of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via agreements D2019-1908080004 and D2019-1906200003.

Mention of commercial products does not imply endorsement by the authors or their institutions.

## References

- [1] M. E. Davison, "The ill-conditioned nature of the limited angle tomography problem," *SIAM Journal on Applied Mathematics*, vol. 43, no. 2, pp. 428–448, 1983.
- [2] M. Rantala, S. Vanska, S. Jarvenpaa, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen, "Wavelet-based reconstruction for limited-angle X-ray tomography," *IEEE Transactions on Medical Imaging*, vol. 25, no. 2, pp. 210–217, 2006.
- [3] P. J. S. Ferreira, "Interpolation and the discrete Papoulis-Gerchberg algorithm," *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2596–2606, 1994.
- [4] Y. Huang, X. Huang, O. Taubmann, Y. Xia, V. Haase, J. Hornegger, G. Lauritsch, and A. Maier, "Restoration of missing data in limited angle tomography based on Helgason–Ludwig consistency conditions," *Biomedical Physics & Engineering Express*, vol. 3, no. 3, p. 035015, 2017.
- [5] Z. H. Levine, T. J. Blattner, A. P. Peskin, and A. L. Pinter, "Scatter corrections in X-ray computed tomography: A physics-based analysis," *J. Res. of the Natl. Inst. of Stand. and Tech.*, vol. 124, p. 124013, 2019.
- [6] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Transactions on Image Processing*, vol. 2, pp. 296–310, 1993.
- [7] T. Sato, S. J. Norton, M. Linzer, O. Ikeda, and M. Hirama, "Tomographic image reconstruction from limited projections using iterative revisions in image and transform spaces," *Applied Optics*, vol. 20, no. 3, pp. 395–399, 1981.
- [8] D. Verhoeven, "Limited-data computed tomography algorithms for the physical sciences," *Applied Optics*, vol. 32, no. 20, pp. 3736–3754, 1993.
- [9] Y. Huang, Y. Lu, O. Taubmann, G. Lauritsch, and A. Maier, "Traditional machine learning for limited angle tomography," *International journal of computer assisted radiology and surgery*, vol. 14, no. 1, pp. 11–19, 2019.
- [10] T. Würfl, M. Hoffmann, V. Christlein, K. Breininger, Y. Huang, M. Unberath, and A. K. Maier, "Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1454–1463, 2018.
- [11] Y. Huang, S. Wang, Y. Guan, and A. Maier, "Limited angle tomography for transmission X-ray microscopy using deep learning," *Journal of Synchrotron Radiation*, vol. 27, no. 2, pp. 477–485, 2020.
- [12] T. A. Bubba, G. Kutyniok, M. Lassas, M. März, W. Samek, S. Siltanen, and V. Srinivasan, "Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography," *Inverse Problems*, vol. 35, no. 6, p. 064002, 2019.
- [13] Y. Huang, A. Preuhs, G. Lauritsch, M. Manhart, X. Huang, and A. Maier, "Data consistent artifact reduction for limited angle tomography with deep learning prior," in *International Workshop on Machine Learning for Medical Image Reconstruction*, pp. 101–112, Springer, 2019.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: convolutional networks for biomedical image segmentation," in *Medical Image Com-*

puting and Computer-Assisted Intervention (MICCAI), pp. 234–241, LNCS 9351, 2015.

- [15] P. Szypryt, D. A. Bennett, W. J. Boone, A. L. Dagel, G. Dalton, W. B. Doriese, M. Durkin, J. W. Fowler, E. J. Garboczi, J. D. Gard, G. C. Hilton, J. Imrek, E. S. Jimenez, V. Y. Kotsubo, K. Larson, Z. H. Levine, J. A. B. Mates, D. McArthur, K. M. Morgan, N. Nakamura, G. C. O’Neil, N. J. Ortiz, C. G. Pappas, C. D. Reintsema, D. R. Schmidt, D. S. Swetz, K. R. Thompson, J. N. Ullom, C. Walker, J. C. Weber, A. L. Wessels, and J. W. Wheeler, “Design of a 3000 pixel transition-edge sensor X-ray spectrometer for microcircuit tomography,” *IEEE Trans. Appl. Superconductivity*, vol. 31, pp. 1–5, 2021.
- [16] M. Mirza and S. Osindero, “Conditional generative adversarial nets,” *arXiv:1411.1784*, 2014.
- [17] M. Holden, M. Pereyra, and K. C. Zygalakis, “Bayesian imaging with data-driven priors encoded by neural networks: Theory, methods, and algorithms,” *arXiv:2103.10182*, 2021.
- [18] S. P. Nolan, L. Pezzè, and A. Smerzi, “Frequentist parameter estimation with supervised learning,” *arXiv:2105.12302*, 2021.
- [19] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, “Spectral normalization for generative adversarial networks,” *arXiv:1802.05957*, 2018.
- [20] H. Wang, Y. Zhu, B. Green, H. Adam, A. Yuille, and L.-C. Chen, “Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation,” in *European Conference on Computer Vision*, pp. 108–126, Springer, 2020.
- [21] J. Andén and S. Mallat, “Deep scattering spectrum,” *IEEE Transactions on Signal Processing*, vol. 62, no. 16, pp. 4114–4128, 2014.
- [22] S. Mallat, “Group invariant scattering,” *Communications on Pure and Applied Mathematics*, vol. 65, no. 10, pp. 1331–1398, 2012.
- [23] H. De Vries, F. Strub, J. Mary, H. Larochelle, O. Pietquin, and A. Courville, “Modulating early visual processing by language,” *arXiv:1707.00683*, 2017.
- [24] A. Reuther, J. Kepner, C. Byun, S. Samsi, W. Arcand, D. Bestor, B. Bergeron, V. Gadepally, M. Houle, M. Hubbell, M. Jones, A. Klein, L. Milechin, J. Mullen, A. Prout, A. Rosa, C. Yee, and P. Michaleas, “Interactive supercomputing on 40,000 cores for machine learning and data analysis,” in *2018 IEEE High Performance Extreme Computing Conference (HPEC)*, pp. 1–6, IEEE, 2018.
- [25] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv:1412.6980*, 2014.
- [26] A. Radford, L. Metz, and S. Chintala, “Unsupervised representation learning with deep convolutional generative adversarial networks,” *arXiv:1511.06434*, 2015.
- [27] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter, “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.

## Author Biography

ZG is a graduate student at MIT EECS, who is working on computational imaging and machine learning. GB is a fellow of Optica (formerly known as Optical Society of America) and has been on the MIT faculty since 1999. His interests are in the intersection of computational imaging and machine learning, explicitly embedding imaging physics. ZHL is a Fellow of the American Physical Society and has been a Physicist at NIST since 1994. He led the first team which published a reconstruction of an integrated circuit interconnect. MEG is physicist at Sandia National Laboratories working on magneto-inertial fusion. He has worked

at LLNL, on advanced ICF concepts and biomedical device design, and in the petroleum and mining industry on the use of machine learning.