Data-efficient Machine Learning for Computational Imaging

by

Gavin (Zhen) Guo

B.A., University of California, Berkeley (2018)
S.M. Massachusetts Institute of Technology (2022)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 2024

Authored by:	Gavin (Zhen) Guo Department of Electrical Engineering and Computer Science May 17, 2024
Certified by:	George Barbastathis Professor of Mechanical Engineering, Thesis Supervisor
Accepted by:	Leslie A. Kolodziejski Professor of Electrical Engineering and Computer Science Chair, Department Committee on Graduate Students

Data-efficient Machine Learning for Computational Imaging

by

Gavin (Zhen) Guo

Submitted to the Department of Electrical Engineering and Computer Science
on May 17, 2024 in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

ABSTRACT

This thesis presents a method that improves data efficiency in computational imaging by incorporating prior knowledge from physical models into machine learning algorithms. Our approach optimizes image reconstruction from sparse and noisy datasets by utilizing physical constraints to guide deep learning models. This integration accelerates the imaging workflow, minimizes the need for large datasets, and improves resilience to measurement noise. The key insight is that physical model-based priors can regularize deep learning for more robust performance. Experiments demonstrate how this physics-assisted machine learning technique enables faster, more accurate, and reliable imaging. By facilitating high-quality imaging from limited data, this method has the potential to advance applications in healthcare, material studies, and industrial inspection. One of the highlights of our method is the application of real-time 2D imaging for improving 3D printing. High-performance manufacturing is achieved by training a neural model combined with a system of dynamic equations. The thesis offers a framework that seamlessly integrates physical insights and data-driven methods, enabling advances beyond what either approach could achieve alone.

Thesis supervisor: George Barbastathis
Title: Professor of Mechanical Engineering

Acknowledgments

I would like to extend my deepest gratitude to a number of individuals whose support and guidance have been invaluable throughout this journey. First and foremost, my heartfelt thanks go to my parents, whose unwavering support and encouragement have been my foundation and strength. Their belief in me has been a constant source of motivation and has made all the difference.

I am also incredibly grateful to my fiancée, whose companionship, understanding, and love have been my sanctuary. Her presence has been a source of joy and comfort, reminding me of the importance of balance and happiness outside the academic world.

Special thanks are due to my advisor, Professor George, for his exemplary guidance and mentorship. His insights, expertise, and patience have not only shaped this work but have also profoundly influenced my growth as a researcher and thinker.

I cannot forget to acknowledge all my collaborators and group members, whose collective efforts and dedication have contributed significantly to our projects. The spirit of teamwork and camaraderie within our group has been truly inspiring, and I am thankful for the opportunity to work alongside such talented and motivated individuals.

To each and every one of you, thank you for your role in this journey. Your support has been a cornerstone of my achievements, and I am deeply appreciative of your contributions.

Biographical Sketch

Gavin (Zhen) Guo was born on May 25, 1993 in Shandong, China. He pursued his passion for physics at the University of California, Berkeley, where he earned his Bachelor of Arts degree in 2018. Following his undergraduate studies, Gavin worked as a research engineer at UC Berkeley's College of Engineering for a year. In 2019, he enrolled in the Department of Electrical Engineering and Computer Science at the Massachusetts Institute of Technology to pursue graduate studies under the guidance of Professor George Barbastathis in the 3D Optics Group.

Alongside his academic endeavors, Gavin has co-founded several technology startups and a venture Decentralized Autonomous Organization (DAO), driven by his desire to effect positive change through innovation. His journey reflects a strong commitment to applying technology to develop practical and impactful solutions to real-world problems.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Computational imaging is a process that indirectly reconstructs images from measurements using computational algorithms. It covers a wide range of applications, including computational microscopy [1], tomographic imaging [2], MRI (Magnetic resonance imaging) [3], ultrasound imaging [4], and more. By integrating sensing and computation, computational imaging systems can access information that would otherwise be inaccessible. These techniques play a crucial role in various fields, from medical diagnostics to materials science, by revealing the internal structures of objects [5, 6, 7, 8, 9, 10, 11, 12].

However, achieving high-quality imaging often requires extensive data acquisition and computational efforts, which can limit the temporal resolution and overall throughput of imaging systems. This challenge stems from the ill-posedness and ill-conditioning of the inverse problem, where limited-angle, sparse, and noisy measurements result in deficits in the Fourier-space information [13, 14, 15], leading to suboptimal results with conventional algorithms.

Model-based reconstruction strategies have attempted to address these issues by incorporating prior knowledge about the target object using iterative optimization with prior terms. However, these approaches are computationally demanding for large datasets, and selecting ideal priors and their weights often involves trial and error. As an alternative, machine learning algorithms, particularly deep neural networks, offer the ability to learn from data distributions and establish direct correlations between projections and reconstructions, inherently encoding priors about object properties and noise for regularization [16, 17, 18, 19, 20, 21]. Nevertheless, their reliance on large datasets can lead to inaccuracies and instabilities when training distributions are narrow, especially when applied to real-world data that may differ from the training set [22, 23]. Physics-assisted machine learning presents a promising solution to these limitations by integrating physics-based models with the learning capabilities of neural networks [24, 25, 26, 27, 28]. This approach ensures adherence to real-world physics while utilizing learned priors about object characteristics and noise from training data. By doing so, physics-assisted machine learning improves data efficiency, reconstruction speed, and resilience to noise, particularly under stringent data constraints often encountered in imaging applications where collecting large datasets is difficult.

In this thesis, we demonstrate the effectiveness of physics-assisted machine learning, also referred to as data-efficient machine learning, in overcoming existing limitations and enhancing imaging accuracy and efficiency. By combining physics-based models and learned priors,

data-efficient machine learning can reduce the amount of data required for training, making it an attractive solution for applications where data acquisition is challenging or expensive. This approach has the potential to advance fields such as medical diagnostics, materials science, and additive manufacturing by enabling more efficient and accurate computational imaging techniques.

1.1 General framework for computational imaging

Computational imaging is a powerful approach that enables the reconstruction of objects that are difficult to image directly. This includes objects that are too small, reaching the diffraction limit, transparent, hard to see with the naked eye, or have hidden internal structures. The general framework for computational imaging can be succinctly described by the following equation:

$$g = H(f), \tag{1.1}$$

where g represents the measured image data, f denotes the object being imaged, and H is the imaging system's response, often termed as the point spread function (PSF) or the system's transfer function. This equation encapsulates the essence of computational imaging by modeling how an object of interest is transformed by the imaging system to produce the observed data. In situations where direct imaging is challenging, the goal is to indirectly measure what is accessible and then computationally reconstruct the desired object.

Depending on the imaging modality in question, H can represent a wide array of physical processes. For example, in optical imaging, H could represent the diffraction and interference of light as it passes through the imaging system. In magnetic resonance imaging (MRI), H would model the spatial encoding of magnetic fields. Therefore, understanding and accurately modeling H is crucial, as it allows for the manipulation of g to extract detailed information about f.

In practice, the imaging system is also subject to noise from the environment and measuring instruments. Two key sources of noise are shot noise due to the quantum nature of light, and thermal noise arising from fluctuations in the electronic circuitry. Shot noise follows Poisson statistics, while thermal noise approximately follows Gaussian statistics. The combination of these two noise sources leads to a hybrid probability distribution for the measured signal at each pixel [29].

1.1.1 Shot Noise

Discrete photons arrive at the detector at random times, following a Poisson distribution. The probability of pixel m receiving k photons, given an expected photon count $\tilde{g}_m$ is:

$$p(g_m = k|\tilde{g}_m) = \exp(-\tilde{g}_m)\frac{\tilde{g}_m^k}{k!}. \tag{1.2}$$

The mean and variance of this distribution are both equal to the expected photon count $\tilde{g}_m$. The pixel-wise signal-to-noise ratio (SNR) due to shot noise is:

$$(\text{SNR})\text{pix} = \sqrt{\tilde{g}_m}. \tag{1.3}$$

Thus, the relative noise decreases with the square root of the photon flux. Shot noise has a signal-dependent nature, with bright pixels being less noisy than dark pixels.

1.1.2 Thermal Noise

Thermal noise arises from charge carrier fluctuations in the electronic circuitry and approximately follows a Gaussian distribution with zero mean and variance σ^2 dependent on temperature T, resistance R, and bandwidth $\Delta\nu$:

$$\sigma^2 \approx \frac{4k_B T \Delta\nu}{R}, \tag{1.4}$$

where k_B is Boltzmann's constant. The probability density of a measured pixel value g_m is:

$$p(g_m|\tilde{g}_m) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(g_m - \tilde{g}_m)^2}{2\sigma^2}\right). \tag{1.5}$$

Thermal noise is additive and signal-independent. The pixel-wise SNR is simply $\tilde{g}_m/\sigma$.

1.1.3 Hybrid noise statistics

In realistic imaging systems, both shot and thermal noise are present. The probability density of the measured signal considering both noise sources is approximately:

$$p(g_m|\tilde{g}_m) \approx \frac{1}{\sqrt{2\pi(\tilde{g}_m + \sigma^2)}} \exp\left(-\frac{(g_m - \tilde{g}_m)^2}{2(\tilde{g}_m + \sigma^2)}\right). \tag{1.6}$$

This is a Gaussian distribution with signal-dependent variance $\tilde{g}_m + \sigma^2$. The combination of shot and thermal noise leads to a noise profile that is worse than either noise source alone, particularly at low light levels.

Assuming Poisson statistics in the detection system, the forward model is modified as

$$g^* \sim \mathscr{P}(g), \tag{1.7}$$

where g^* is the noisy measurements vector, and $\mathscr{P}$ is a vector of Poisson distributions whose parameters are elements of the vector g. The operator $\sim$ means "is drawn from the distribution," following a convention from statistics. Understanding the statistical properties of these noise sources is crucial for developing effective image denoising and reconstruction algorithms. The log-likelihood expressions derived from these probability distributions serve as key ingredients in maximum likelihood and Bayesian estimation approaches to image restoration.

1.2 Inverse estimation of the imaging object

Solving an inverse problem in computational imaging involves estimating the original object from the measured data, given knowledge of the system's response. This requires finding f when g^* and H are known. However, inverse problems are notoriously difficult to solve for three reasons:

1. **Noise and errors:** Measured data are often contaminated with noise, and the system's response may not be perfectly known, making the inverse process more challenging.

2. **Ill-conditioning:** Ill-conditioning occurs when small perturbations in the measured data, such as noise, lead to disproportionately large variations in the solution. As a result, even small amounts of noise in the measurements can get significantly amplified, resulting in solutions that deviate greatly from the ground truth.

3. **Ill-posedness:** Ill-posedness deals with more fundamental challenges in inverse problems. It encompasses situations where a unique solution might not exist, or where solutions do not continuously depend on the input data. This often arises from information loss during imaging processes, where the system's response can map different imaging objects to the same measurements.

To solve inverse problems, several approaches are commonly used:

1. **Regularization:** This involves incorporating additional information or constraints into the inversion process to make the problem well-posed. For example, one might assume that the object f is smooth or sparse in a certain domain, which helps in narrowing down the possible solutions [30, 31].

2. **Iterative methods:** These methods start with an initial guess of $\hat{f}$ and iteratively refine it to better fit the measured data. Examples include the gradient descent method [32] and more sophisticated algorithms like the Conjugate Gradient [33] or Expectation Maximization [34].

3. **Machine learning:** Recent approaches use machine learning, particularly deep learning, to learn the approximate mapping to f [35, 24, 36, 37, 38, 39, 40, 41]. These methods typically require training data consisting of pairs of objects and the corresponding measured data, but they can offer fast and accurate reconstructions once trained.

In this thesis, we focus on the third approach, using machine learning for the inverse problem without requiring extensive training data, hence data-efficient machine learning.

1.3 Applying machine learning in computational imaging

Machine learning, particularly deep learning, has emerged as a transformative tool in computational imaging, offering novel ways to address the challenges associated with inverse problems. Deep learning models can be trained to effectively filter out noise from measured data, resulting in more accurate reconstructions compared to traditional methods [42, 43, 20]. These models can learn the inverse mapping for a wide range of imaging systems without requiring explicit mathematical formulations, making them adaptable to different imaging modalities. Once trained, deep learning models can perform reconstructions much faster than iterative algorithms, as the computation is often a single forward pass, which is critical for real-time applications.

Various implementation strategies have been explored in recent years. One of the pioneering works in this field is the use of convolutional neural networks (CNNs) for end-to-end learning in computational imaging. Krizhevsky et al. [44] first showcased the effectiveness of CNNs in image classification tasks, paving the way for their application in inverse problems. CNNs have then been successfully applied to inverse problems in lensless imaging [35], compressed sensing MRI [45], and holographic phase retrieval [46]. Generative models, such as generative adversarial networks (GANs) and variational autoencoders (VAEs), have also been explored for computational imaging. Yang et al. [47] used a GAN-based approach for learning the inverse mapping in compressive sensing, while Zhang et al. [38] employed cycleGAN [48] to address the phase problem in real-time when no phase reconstructions but good simulations or data from other experiments are available. VAEs have been used by Francesco et al. [49] for reconstructions in holographic imaging and imaging through scattering media. A neural network based on deep image prior [50, 51] can simultaneously learn the illumination spectrum and recover the complex object from a single diffraction measurement in digital holography [52]. For imaging through scattering media, a recurrent neural network (RNN) approach can exploit spatiotemporal dynamics to reveal hidden phase objects [53]. Beyond phase retrieval, treating a sequence of images from different illumination angles as a dynamical system allowed an RNN to reconstruct object interiors with limited-angle measurements under both weak and strong scattering [40].

Deep learning has been applied to various imaging modalities, including holography and coherent imaging [54, 55, 56, 57, 58, 26], imaging through scattering media and diffraction tomography [59, 60, 61, 53], ptychographic image reconstruction [62, 63, 64, 65], model-based nonlinear inverse scattering [66], optical microscopy [67, 68], and uncertainty quantification in phase imaging [69]. These works demonstrate the potential of deep learning in improving reconstruction quality, speed, and enabling new capabilities in computational imaging. In the field of holography and coherent imaging, deep learning has enabled new image reconstruction and phase recovery techniques that are faster and more robust compared to traditional iterative methods. CNNs can be trained to improve reconstruction quality [54, 56], and compute the computer-generated holograms non-iteratively [55, 58]. Novel deep learning methods have been proposed for ptychographic image reconstruction, such as PtychoDV [70], which consists of a vision transformer [71] for initial image generation and a deep unrolling network [72, 73] for refinement. PtychoDV outperforms existing deep learning methods and achieves results competitive with iterative algorithms while being substantially faster. Plug-and-play priors (PPP) have been used in combination with deep learning for model-based nonlinear inverse scattering problems in imaging. A FISTA (fast iterative shrinkage-thresholding algorithm) variant of PPP has been proposed, which enables advanced denoising priors for regularization [66]. Online plug-and-play algorithms, such as PnP-SPGM (plug-and-play stochastic proximal gradient method), have been developed for regularized image reconstruction from a large number of measurements, making them scalable to large datasets [74]. Deep learning has also been used to enhance optical microscopy images, improving spatial resolution, field of view (FOV), and depth of field (DOF) without requiring hardware modifications. CNNs trained on pairs of low-resolution and high-resolution microscopic images can successfully enhance resolution, FOV, and DOF of test images not used in training, demonstrating generalizability across different tissue types and staining methods [67, 68]. In addition, uncertainty quantification in deep learning-based phase imaging

has been explored using Bayesian convolutional neural networks. The combination of multiplexed illumination with uncertainty-quantifying deep learning enables reliable and scalable high-space-bandwidth product phase imaging, with the uncertainty predictions providing a reliability measure for deep learning predictions in scientific imaging applications [69].

Deep learning for computational imaging often involves intelligent representation and fusion of spatial frequencies. For example, separately processing low and high frequencies with dedicated deep neural networks (DNNs) before optimally synthesizing them has enabled robust, high-quality phase retrieval [37]. Physics-assisted DNNs incorporate elements of the physical model, such as an initial approximant, and have demonstrated superior performance to end-to-end learning approaches [75, 24, 21, 40]. These physics-assisted approaches differ from physics-informed neural networks [76, 77] by utilizing conventional algorithms to generate inputs to the neural network that already satisfy known physical constraints. By separating physics from network optimization, physics-assisted strategy allows the network to focus on learning priors from datasets like ImageNet [78], which has been empirically shown to improve results in recent research [39].

1.3.1 Limitations and the need for data-efficient machine learning

While deep learning has shown great promise in computational imaging, prior works have several limitations that need to be addressed. One of the main challenges is the generalization ability of the learned models to unseen data, particularly when the imaging conditions or noise levels differ from those encountered during training. Many existing methods rely on supervised learning with large training datasets [35, 24, 26, 37, 40, 41], raising concerns regarding the instability and lack of generalizability in deep learning methods for image reconstruction [22, 79].

Prior works show that deep learning approaches have some level of generalization ability for unseen test data. Sinha et al. [35] demonstrate that a DNN is robust to moderate perturbations in sensor displacement and exhibits some degree of shift and rotation invariance when applied to test inputs. More importantly, a DNN trained on face or natural images can reconstruct test images from completely different classes, such as handwritten digits and characters, indicating some level of generalization ability. However, the choice of training dataset strongly impacts the cross-domain generalization performance of the trained neural network. Deng et al. [39] show that training on a higher entropy, more generic dataset like ImageNet leads to better generalization to unseen classes compared to training on a lower entropy, more constrained dataset like MNIST [80]. Kang et al. [40] analyze cross-domain generalization in terms of sparsity and find that networks perform better when trained and tested on sparse samples compared to dense samples.

In a related study, Kang et al. [26] investigate the generalizability of DNNs for low photon phase imaging and find that it depends on the strength of the priors in the training data. Strong priors enable the DNN to handle high noise even without coherent modulation imaging (CMI), a technique that introduces a physical constraint by applying random phase modulation to the optical field diffracted from the object at an intermediate distance between the object and the camera. The CMI scheme effectively improves ill-posedness and eliminates ambiguous solutions in the inverse estimate of the phase. Kang et al. show that weak priors benefit more from the CMI scheme in combination with the DNN to improve recon-

structions under severe noise. Wu et al. [79] propose the Analytic Compressed Iterative Deep (ACID) framework, which synergizes a deep reconstruction network, kernel awareness from compressed sensing inspired processing, and iterative refinement to minimize data residuals. ACID is shown to be resilient against adversarial attacks and eliminates instabilities such as strong artifacts from tiny perturbations, missing small features, and decreased performance with increased input data.

Another important consideration in the generalization ability of deep learning methods is the assumption of comparable noise statistics between training and test datasets, which was made in previous studies [24, 26, 41]. For measurements that are corrupted by Poisson noise, the reconstruction quality from using the learned prior is often evaluated by test data with the same number of photons per ray as the training data. However, in practical systems, training and test data might have different noise levels, leading to out-of-distribution data and degradation in reconstruction fidelity [22, 79]. To mitigate these generalization issues, one approach is to train models using a diverse range of datasets that span various noise levels and imaging conditions. However, this is often impractical due to the long acquisition times required to collect such extensive training data. Imaging systems, especially in scientific and medical applications, may have limited throughput, making it challenging to acquire large datasets that cover all possible scenarios. Some prior works have explored techniques like transfer learning to address these limitations. Kang et al. [41] demonstrate that transfer learning can be utilized to reduce the amount of training data required when working with new integrated circuits (ICs) that have different design rules at low photon phase imaging, allowing the network to adapt to variations in the imaging objects. However, there is still a need for more effective data-efficient learning strategies.

Data-efficient machine learning aims to build upon the limitations of prior works by developing strategies that incorporate prior knowledge and generative modeling to enable robust and adaptable models for high-quality reconstructions in data-limited scenarios. By applying prior knowledge about the imaging system and the objects being imaged, such as physical laws, geometrical constraints, spatial correlations, or noise statistics, data-efficient learning guides the models towards more plausible reconstructions even with limited training samples. This is in contrast to many prior works that rely solely on the dataset itself to learn the reconstruction mapping.

1.3.2 Contributions of this thesis

This thesis focuses on developing and applying data-efficient machine learning techniques in computational imaging to address the limitations of traditional deep learning approaches. The main contributions are:

1. We propose data-efficient machine learning frameworks that incorporate prior knowledge about the imaging system and objects into the learning process, enabling high-quality reconstructions with limited training data. These frameworks utilize physics-assisted neural networks and generative modeling to enhance robustness and generalization, without making complicated neural model architectures.

2. We demonstrate the effectiveness of our data-efficient machine learning approaches through experiments on tomographic reconstruction and imaging 3D printed objects.

Results show that our methods consistently outperform traditional deep learning approaches and the state-of-the-art iterative reconstruction algorithms, especially with limited or noisy training data.

3. We provide a comprehensive analysis of the performance and generalization capabilities of our data-efficient machine learning frameworks, investigating their robustness to different noise levels, imaging conditions, and data distributions. This analysis offers valuable insights into the practical applicability and limitations of these approaches in real-world computational imaging tasks.

We have released open-source implementations of our data-efficient machine learning frameworks and datasets in an effort to promote reproducibility and encourage the research community to build upon our works. By developing and applying data-efficient machine learning techniques in computational imaging, this thesis expands the possibilities in this field, enabling high-quality reconstructions in data-limited scenarios. The proposed methods and insights have the potential to impact various applications, such as medical diagnostics, materials science, and biological research, by enabling faster, more accurate, and more reliable imaging capabilities.

The following chapters provide detailed descriptions of our data-efficient machine learning frameworks, their implementation strategies, and their application to computational imaging tasks. Chapter 2 introduces our Physics-assisted Generative Adversarial Network (PGAN) for X-ray tomography, incorporating physical priors for efficient learning. We showcase PGAN's performance in achieving high-quality reconstructions with limited projection angles and reduced photon requirements. Chapter 3 discusses our noise-resilient deep learning framework for integrated circuit tomography, combining Gaussian noise and sparsity-promoting priors into MAP reconstructions before sending to the neural network. We demonstrate improved noise resilience of the network without requiring additional training data with varying noise levels. Chapter 4 presents our data-efficient neural modeling approach for high-precision 3D printing using two-color projection micro-stereolithography (PμSL). We introduce our neural network model, integrate real-time diffractive imaging and post-printing microscope imaging, and employ a two-stage training approach. We demonstrate strong performance in predicting final printed patterns with limited training data. Final conclusions are in chapter 5. These chapters collectively showcase the effectiveness of our data-efficient machine learning frameworks in addressing computational imaging challenges in data-limited scenarios.

Chapter 2

Physics-assisted generative adversarial network for X-ray tomography

X-ray tomography is a non-invasive 3D imaging technique with applications in biomedical imaging, materials science, and electronic inspection. It involves acquiring 2D radiographs from various angles and reconstructing them into 3D objects using computed tomography algorithms [81, 82, 83, 84]. The reconstruction process in X-ray tomography is often an ill-conditioned and ill-posed inverse problem due to inadequate sampling of high frequencies by discrete angular measurements. Practical constraints, such as limited-angle measurements and low photon counts in radiation-sensitive samples, exacerbate this challenge [85].

Direct reconstruction algorithms such as filtered back-projection (FBP) are often inadequate as they can generate reconstructions with noise and streak artifacts [86]. Iterative algorithms whose objective function includes a term representing prior knowledge about the object may compensate for the deficits in Fourier space coverage and often produce higher fidelity results [87, 88, 89]. When prior knowledge is used in an iterative algorithm, the optimization balances minimization of the residual of the simulated measurements from a reconstructed object against minimization of the regularization term. Assumed priors such as sparsity, total variation, and nonlocal similarity priors have been used in X-ray tomography [90, 91, 92]. However, without trial and error, it is not straightforward to choose the appropriate prior and regularization weight for a given set of objects. A prior distribution may also be learned from the dataset itself by a machine learning algorithm. Using a large amount of paired training data, a prior can be determined through exploring the statistical properties of the training distributions, improving the reconstruction quality. Recently, learned priors have been successfully applied to tomography in treating the ill-conditioned inverse problem. In particular, deep learning, a subset of machine learning that is based on artificial neural networks, achieved promising results [16, 17, 18, 19, 20, 21]. For example, efforts have been made in using learned priors from deep neural networks to recover boundary information [18], and to generate missing projections with a data-consistent reconstruction method [19]. However, reports have shown that these methods suffer from reconstruction artifacts and instabilities [22, 23]. To avoid these issues, some works use reconstructions from a direct or iterative algorithm [75, 93, 20, 21, 41], or use a two-step deep learning strategy to generate reconstructions that are empirically more stable and accurate [79].

In this chapter, we introduce a Physics-assisted Generative Adversarial Network (PGAN)

for limited-angle, low-photon X-ray tomography of 3D integrated circuits. PGAN addresses the key challenges of the inverse problem in the following ways:

1. The generative model in PGAN combines known physics with deep learning to learn an appropriate prior from the dataset, eliminating the need for manual trial and error in choosing priors and regularization weights for iterative algorithms.

2. By using iterative algorithm-derived maximum likelihood estimates (MLE) as inputs and regularizing reconstructions with the learned prior, PGAN reduces noise and streak artifacts common in conventional reconstruction algorithms like FBP and MLE. Compared to MLE reconstruction, PGAN reduces the photons per ray required to achieve a single error per sample with limited-angle measurements from 5000 to around 500.

3. PGAN outperforms deep learning methods with no or less physics assisting in terms of the photon requirements to achieve a target reconstruction error.

To carry out our numerical explorations, we utilized CircuitFaker, a synthetic circuit generator developed by Zachary Levine from NIST (National Institute of Standards and Technology), to facilitate learning implicit circuit correlations. CircuitFaker plays an important role in providing realistic training data for our PGAN model. Below we will provide an overview of the forward model for X-ray tomography, explore inverse algorithms, present our PGAN development, and demonstrate its effectiveness in enhancing X-ray imaging quality while minimizing photon exposure.

2.1 Forward model for X-ray tomography

An X-ray tomography system typically includes a sample positioned between an X-ray source and a detector. The system captures measurements by rotating the sample through various angles, utilizing a cone-beam geometry to project rays from the source through the object to the detector's center pixel. This conceptual diagram is depicted in Figure 2.1, with the sample being a three-dimensional integrated circuit (IC). Assuming noise-free conditions, the detection model is described by:

$$g^{(0)} = \int dE\, D(E)\, I^{(0)}(E) e^{-\alpha(E) A f}, \tag{2.1}$$

where A represents the system matrix (indicating the distance each ray travels from the source, through the object, to a detector pixel), f is the vector representing the object's voxel compositions, E denotes the photon energy, $\alpha(E)$ is the energy-dependent absorption coefficient, $I^{(0)}(E)$ the initial source intensity, $D(E)$ the detector efficiency, and $g^{(0)}$ the expected photon count for each detector pixel. The exponential function is applied to each component individually. For monochromatic illumination, this model simplifies to:

$$g^{(0)} = N_0 e^{-\alpha A f}, \tag{2.2}$$

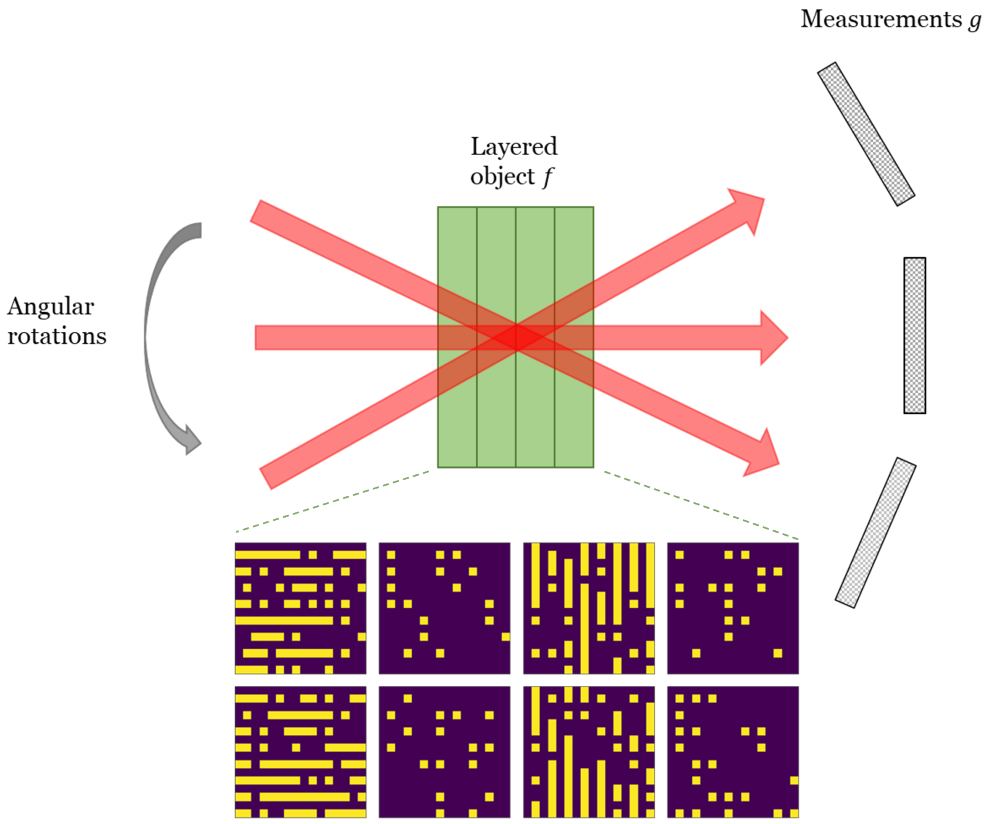


Figure 2.1: A conceptual diagram for our imaging system (IC as the object).

with N_0 being the expected photon count per ray, A the system matrix as defined previously, and α the absorption coefficient. This relationship can be reformulated as a linear equation:

$$\ln g^{(0)} - \ln N_0 = -\alpha A f, \tag{2.3}$$

applying the natural logarithm component-wise. Our forward model simulations account for the Poisson distribution of measured photon counts.

It is important to note that the entire chapter's analysis is subject to the Radon approximation, which assumes that the X-rays travel along straight lines through the object and that the scattering of X-rays is negligible. This approximation is widely used in X-ray tomography due to its simplicity and computational efficiency. However, in cases where scattering effects are significant, such as in strongly scattering materials or at low X-ray energies, a more complex forward model that accounts for scattering may be necessary. In such scenarios, the analysis presented in this chapter may not be directly applicable, and further modifications to the forward model and reconstruction algorithms would be required to address the scattering effects accurately.

2.2 Inverse algorithms for X-ray tomography

2.2.1 Filtered back-projection

Filtered back-projection (FBP) has long been a standard reconstruction algorithm in X-ray tomography, performing a two-step process that first applies a filter to the measurements and then back-projects them to construct the object image. However, FBP's performance is notably affected by the sampling rate and noise levels in the Fourier space, leading to unsatisfactory reconstructions with noise and streak artifacts under limited-angle and low-photon conditions [86].

2.2.2 Iterative algorithms with prior regularizer

Contrastingly, iterative algorithms enhanced with prior regularization, such as the Wiener-Tikhonov method, offer improvements by optimizing:

$$\hat{f} = \arg\min_{f} \left\{ \| - \alpha A f - \ln g + \ln N_0 \|^2 + \beta \, \Psi(f) \right\}, \tag{2.4}$$

where $\hat{f}$ represents the reconstruction, $\|\cdot\|$ denotes the L^2 norm, $\Psi(f)$ the regularization term or Bayesian prior, and β the regularization parameter. This approach, assuming Gaussian noise, begins with an initial object guess, simulates measurements, compares these with actual measurements, and iteratively updates the object model to minimize discrepancies, including those from prior assumptions. This iterative refinement, driven by the choice of $\Psi(f)$, aims to suppress artifacts and preserve edges. However, acquiring accurate prior information can be challenging, and determining the optimal regularization parameter often requires a trial-and-error approach.

2.2.3 Deep reconstruction networks with learned prior

Deep-learning-based inversion, using deep reconstruction networks with learned priors, has emerged as an alternative approach in X-ray tomography. These networks employ learned prior distributions, derived from supervised training on datasets of ground truth objects and their corresponding measurements, to generate high-fidelity reconstructions.

There are two primary categories of deep reconstruction networks utilizing learned priors. The first is the End-to-End model, which establishes a direct correlation between measurements and object reconstructions using paired datasets for training. This model operates without explicitly incorporating physical models, attempting to implicitly learn both the inverse physics and object priors. However, End-to-End approaches have been criticized for their potential to produce unstable reconstructions due to ambiguities in the system matrix's null space [22, 23].

To address these challenges, a second type of network incorporates some degree of physics into the reconstruction process. Initially, an FBP algorithm generates a preliminary, though noisy, reconstruction from the measurements. Subsequently, a deep network refines this initial image by eliminating noise and artifacts [75, 93, 20, 21, 41]. This approach allows the network to focus on learning the object prior from the FBP output and ground truth,

bypassing the need to infer the inverse physical process. For enhanced stability, some studies have replaced FBP with another deep network, aiming to mitigate reconstruction instabilities more effectively [94, 79].

2.3 Physics-assisted Generative Adversarial Network

Our Physics-assisted Generative Adversarial Network (PGAN) advances deep reconstruction networks by employing a two-step process for image reconstruction. Initially, instead of using an FBP approach, we apply a maximum-likelihood estimate derived from an iterative algorithm that integrates known physical principles. This first step includes the forward imaging geometry of X-ray tomography and accounts for Poisson noise characteristic of low-photon measurements. Subsequently, a generative model refines this estimate using a learned prior $\Psi(f)$, thereby enhancing the quality of the reconstruction. PGAN thus combines the precision of physics-based iterative algorithms with the adaptability of deep learning through learned priors, offering improvements in reconstruction accuracy.

2.3.1 Maximum-likelihood estimate

Maximum-likelihood estimates for tomographic reconstructions are achieved by optimizing an objective function that reflects the projective geometry and Poisson statistics of the imaging process. These estimates not only provide inputs for our generative models but also serve as the baseline for evaluating our machine learning approaches. The target is to optimize $\tilde{f}$, the reconstruction from tomographic measurements g (photon counts per detector pixel), under the assumption that measurement noise follows a Poisson distribution. The optimization problem is defined as:

$$\tilde{f}(g) = \arg\max_{f^{(0)}} \left[L_{\mathrm{MLE}}(g|f^{(0)}) + \Psi(f^{(0)}) \right] \quad \text{and} \tag{2.5}$$

$$L_{\mathrm{MLE}}(g|f^{(0)}) = -\sum_i \left[\ln g_i! - g_i \ln g_i^{(0)} + g_i^{(0)} \right]. \tag{2.6}$$

where L_{MLE} represents the log-likelihood based on Poisson statistics, Ψ denotes a regularization function or the log of the Bayesian prior, $g^{(0)}$ simulates the measurement from a proposed object $f^{(0)}$ as per Eq. 2.2 without considering noise, $\sum_i$ aggregates over all detector pixel measurements at various angles, and $\tilde{f}$ is the reconstruction maximizing the log-likelihood. In our implementation, the projective reconstruction Fortran 95 code using maximum likelihood method used here has been presented recently as parts of a study on scatter corrections in tomography [95] and another on diffractive tomography [96]. The key point is that the maximum likelihood objective function as formulated by Sauer and Bouman [97] is minimized using the version of the Broyden-Fletcher-Goldfarb-Shanno algorithm known as L-BFGS-B [98].

2.3.2 Deep generative models

Our approach utilizes a supervised machine learning framework known as the conditional generative adversarial network (cGAN) [99] to enhance 3D reconstructions from X-ray to-

mography under limited projection angles and photon counts. This model uses a prior distribution learned from the data to improve upon reconstructions obtained through maximum-likelihood estimation, which often suffer from artifacts due to incomplete data ("missing cone problem"). The quality of maximum-likelihood reconstructions declines with reduced angular coverage and photon flux, but our cGAN-based model counteracts this degradation, producing outputs that more accurately reflect the true object structure by incorporating the system's imaging geometry and measurement statistics as conditional information.

In its foundational form, a GAN aims to generate data matching a target distribution, utilizing a generator (G) to create examples and a discriminator (D) to evaluate them:

$$\arg\min_{G}\max_{D}\Big(\mathbb{E}_{f}\big[\log D(f)\big]+\mathbb{E}_{z}\big[\log\big(1-D(G(z))\big)\big]\Big), \tag{2.7}$$

where the generator aims to minimize this function while the discriminator aims to maximize it, facilitating a dynamic optimization process. cGAN, an adaptation of GAN, conditions both G and D on additional information, thus providing more control over the generated distribution and addressing the instability issues associated with the original GAN [100]. Our model specifically conditions on the noisy maximum-likelihood estimation $\tilde{f}$, bypassing the need for a random input vector. Training involves pairs of such estimates and their corresponding ground truths, focusing the learning process on improving the fidelity of reconstructions. The discriminator's objective in our cGAN framework is thus modified to:

$$\arg\min_{G}\max_{D}\mathbb{E}_{(f,\tilde{f})}\big[\log D(f)+\log\big(1-D(G(\tilde{f}))\big)\big]. \tag{2.8}$$

Through iterative training, our generative model learns to navigate the complex interplay between generator and discriminator towards achieving a Nash equilibrium, at which point the training is deemed complete. The output of this trained model, denoted as $\hat{f}$, reflects a significantly refined reconstruction, demonstrating the model's capacity to harness physical priors for improved imaging reconstructions.

In our model, the generator is structured as a 3D autoencoder, incorporating an encoder to map the object representation into a latent space, followed by a decoder that reconstructs the object from this latent representation. The discriminator, a 3D convolutional neural network, assesses the quality of the generator's output by predicting the probability that each sample is real or fake, with values closer to 1 indicating a higher likelihood of the sample being real. To ensure training stability, both the generator and discriminator employ spectrally normalized convolutional kernels [101]. The discriminator's role concludes post-training, being unnecessary for testing phases. We explore four deep generative model variants: a baseline, one with axial attention, another incorporating a scattering representation, and a combination of axial attention with scattering representation, each modifying the encoder's design while maintaining consistent decoder and discriminator architectures, detailed further in our Github repo [102].

The baseline model employs cascading 3D convolutional and pooling layers within the encoder to feature-extract from input reconstructions, with an increase in convolutional layers enhancing the encoder's ability to learn complex features [103]. The axial attention variant captures contextual information by integrating full axial attention within the encoder, replacing some 3D convolution layers. This approach decomposes 3D self-attention into sequential

1D attentions along the input's height, width, and depth axes, reducing computational complexity to $\mathcal{O}(hwzm)$ and facilitating global feature extraction beyond the local scope of convolutional kernels [104]. The last two variants incorporate a wavelet scattering transform input, providing multi-scale features without requiring training [105, 106]. This scattering representation, when coupled with renormalization, conditions the generative model to produce more realistic reconstructions by modulating the feature values derived from convolutional layers or axial attention through a fully connected layer transformation [107]. This approach enriches the model with additional features, aiding in the transition from noisy to noiseless reconstructions.

2.4 Evaluation methods

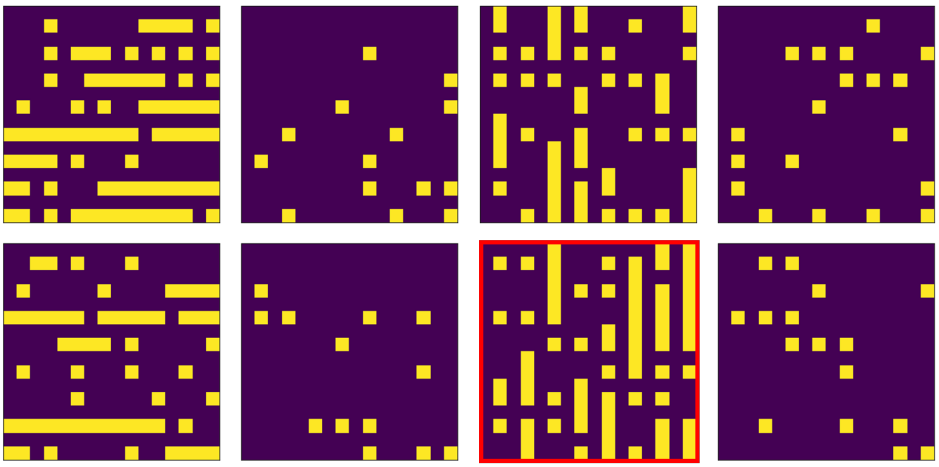


Figure 2.2: Each image is a slice of 2D layer in the z dimension. The value of z increases as a raster scan of the 8 slices shown. Yellow indicates copper and purple indicates silicon. Here, x layers are the first (upper left) and fifth layers (lower left) in z, y layers are the third and seventh layers in z. Others are via layers. The layer highlighted in red is the ground truth circuit layer in the later comparisons.

Our evaluation methods involve two primary components: the generation of synthetic tomographic objects using CircuitFaker and the assessment of reconstruction quality via imaging geometry and bit error rate (BER) analysis.

2.4.1 CircuitFaker for tomographic objects

CircuitFaker is developed by our collaborator Zachary Levine, and it is designed to produce synthetic integrated circuit interconnects by randomly assigning binary voxel values to simulate the circuit's spatial layout. It initializes all voxels to 0 and then populates the circuit using a probabilistic approach to mimic wire seed points and layer types, including x, y, and via layers, with varying probabilities for extending wires (p_w, p_x, p_y, and p_z). For our experiments, we set the circuit dimensions and probabilities as follows: $N_x = N_y = 16$,

$N_z = 8$, $p_w = 0.75$, $p_x = p_y = 0.8$, and $p_z = 0.5$, producing circuits of size $16 \times 16 \times 8$ as shown in Fig. 2.2. The imaging geometry tailored for X-ray tomography of integrated circuits features a voxel size of 0.15 μm $\times$ 0.15 μm $\times$ 0.30 μm, covering a total volume of 2.4 μm^3. The detector, positioned in the x-z plane without tilt ($\varphi = 0°$), captures images at eight tilt angles ranging from $-30°$ to $+22.5°$. We assume a single source point in a cone-beam geometry, neglect minor corrections for source-detector distance variability, and focus solely on copper as the material of interest, with a binary reconstruction outcome based on two primary X-ray lines.

2.4.2 Bit-error-rate formulation

To quantify reconstruction accuracy, we introduce BER as a metric, calculated by:

1. Estimating posterior distributions $p(f_i = 0 \,|\, \tilde{f})$ for each voxel f_i by considering the likelihoods $p(\tilde{f} \,|\, f_i = 0)$ and $p(\tilde{f} \,|\, f_i = 1)$ alongside their priors.

2. Classifying voxels as 0 or 1 based on a threshold derived from the intersection of their distribution likelihoods.

3. Calculating error rates for both binary states by integrating over their misclassified probability density functions.

4. Computing the average BER as $\eta_{\text{avg}} = \eta_0 p_0 + \eta_1 p_1$.

This approach provides a direct measure of the frequency of voxel misclassification, offering insights into the performance of our tomographic reconstructions.

2.5 GAN training

2.5.1 Network architecture

Fig. 2.3 is the detailed network architecture for the deep generative model (the generator). The overall design is based on UNet [108] to perform pixel-by-pixel prediction (for 3D reconstruction, where the 3D object is voxelized by a 3D matrix). The input dimension to the model is in $(16, 16, 8, 1)$. Four DownResBlocks encode the input approximant and produce a latent representation that is in dimension of $(1, 1, 8, 512)$. Four UpResBlocks decode the latent representation to a vector in dimension of $(16, 16, 8, 64)$. Concatenated skip-connections are used in between the last three DownResBlocks and the first three UpResBlocks to preserve high frequency information of the input approximant [39]. Dropout layers are included to prevent over-fitting. The final layer of convolution reduces this vector to a final output in $(16, 16, 8, 1)$, and a Tanh layer forces the final output to the range between -1 and 1. The DownResBlock and UpResBlock share similar topology to the Resblock in ResNet [109], except the use of different 3D sampling layers. Here, we implemented downsampling and upsampling layers that only sample the dimension in height and width but not depth.

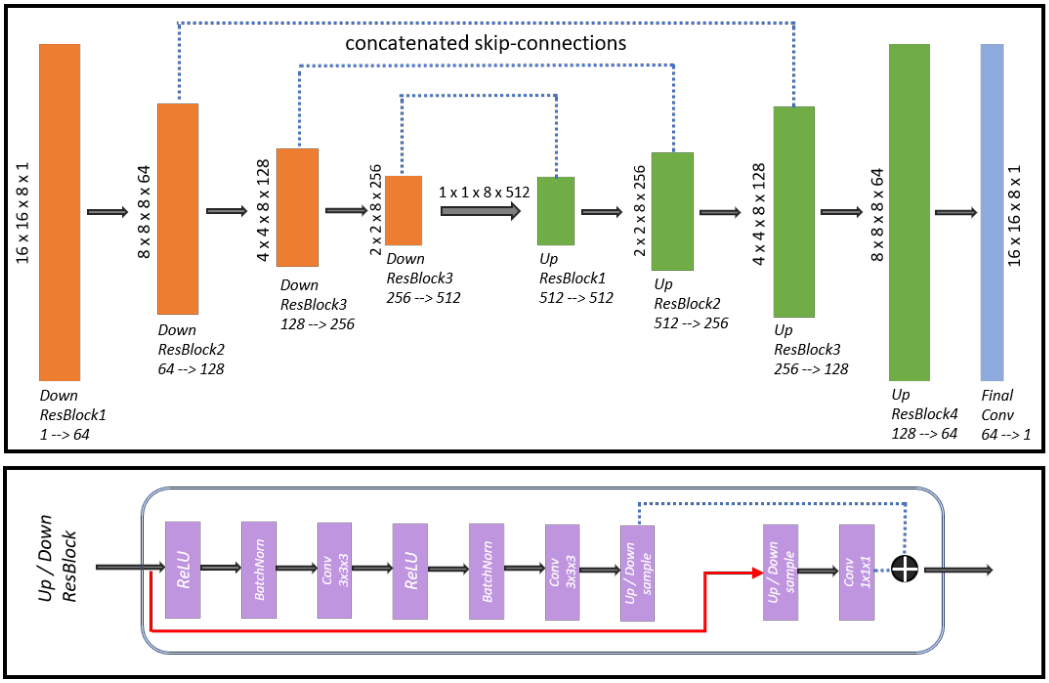


Figure 2.3: Network architecture for the deep generative model (generator).

For the base generative model, feature extraction in the DownResBlock and UpResBlock is achieved by 3D convolutional kernel with spectral normalization [101]. For the axial-attention based model, feature extraction in the DownResBlock is achieved by the mixture of 3D convolutional kernel and axial attention both with spectral normalization.

For models including the wavelet scattering transform, the wavelet representation for the input approximant is first produced by HarmonicScattering3D in the Kymatio package [110] with $J = 2$ (maximum scale of 2^2), integral powers with $\{0.5, 1.0, 2.0, 3.0\}$. Then, the batch normalization layers in the UpResBlock are replaced by conditional batch normalization (CBN) layers [111], where the conditional information is the wavelet representation. Note that the fully connected layers within the CBN are spectrally normalized as well.

The discriminator for all the generative models is the same, with four DownResBlocks bringing the input from dimension $(16, 16, 8)$ to $(2, 2, 1024)$, following with a reduce sum operation to bring it further to a vector of $(1, 1, 1024)$. A fully connected layer followed thereafter to produce a floating point number for classification.

For End-to-End methods, we add an additional DownResBlock at the top of the generator to down sample the measurements from $(32, 32, 8, 1)$ to $(16, 16, 8, 1)$. Concatenated skip-connections are also removed to avoid the issues of overlapping features from two different domains (from projections to object reconstruction).

2.5.2 Training parameters

Our proposed networks are 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 [112]. An Adam optimizer [113] is used with parameters $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The two time-scale

update rule (TTUR) is used to stabilize the training of the generative network [114], where the initial learning rate is 10^{-4} for the generator and 4×10^{-4} for the discriminator. In each iteration, the generator is updated four times while the discriminator is updated once.

Training sets of 1800 reconstructions are generated independently for each condition studied, except that the ground truth is common. The batch size for training is 20. An additional 200 reconstructions per condition are used for testing. The learning rate is reduced by half when the validation loss stops improving for 5 iterations. We set the maximum number of iterations to be 200, and the training stops early when either the validation loss plateaus for 20 iterations, or the minimum learning rate 10^{-8} is reached. This early-stop technique can 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

$$r_{f,\tilde{f}} = \frac{\mathrm{cov}(f, \tilde{f})}{\sigma_f \, \sigma_{\tilde{f}}}, \tag{2.9}$$

where cov is the covariance and σ is the standard deviation. The total objective of training is to find the optimal generator G_{opt} given the approximant $\tilde{f}$ and ground truth f:

$$G_{\mathrm{opt}}(\tilde{f}) = \arg\min_G \max_D \mathbb{E}_{(f,\tilde{f})}\big\{ - r_{f,G(\tilde{f})} + \lambda\big[\log D(f) + \log\big(1 - D(G(\tilde{f}))\big)\big]\big\}. \tag{2.10}$$

The hyper-parameter λ controls the degree of generation from input noise to features. In our experiments, λ ranges from 2^0 to 2^{-6} with an incremental factor of $1/2$. The loss function for GAN is the hinge loss [115], and is defined below:

$$\begin{aligned} L_D &= \mathrm{mean}\big\{\min\{0, 1 - D(f)\}\big\} + \mathrm{mean}\big\{\min\{0, 1 + D(G(\hat{f}))\}\big\}, \\ L_G &= -\mathrm{mean}\{D(G(\hat{f}))\}. \end{aligned} \tag{2.11}$$

Here, L_G is the loss for generator and L_D is the loss for discriminator. The operator min(...) chooses the smaller value between the two inputs. The mean is taken over the batch of the training data.

2.5.3 Convergence and stability of the deep generative network

Initially, Generative Adversarial Networks (GANs) encountered training instability, often leading to model collapse and unsatisfactory outputs. The development of deep convolutional GANs (DCGANs) marked a turning point, introducing structural and training improvements that enhanced GAN stability [116]. Further advancements came with the Wasserstein GAN (WGAN), which offered insights into GANs' control issues and proposed Lipschitz continuity as a solution for improving result quality [117, 118]. Presently, several well-established techniques have been adopted to address GAN training challenges, as we detail in the context of our Physics-assisted Generative Adversarial Network (PGAN).

1. **Spectral Normalization**: Unlike WGAN's gradient clipping or penalty methods for enforcing Lipschitz continuity, spectral normalization achieves this by normalizing the network's weights. This method is both computationally efficient and straightforward to integrate, proving effective across various applications [119, 120, 121].

2. **Hinge Loss**: Coupling hinge loss with spectral normalization has become a standard practice in cutting-edge GANs, noted for its performance enhancement [122].

3. **Two Time-scale Update Rule (TTUR)**: TTUR facilitates the GAN's theoretical convergence to a stable local Nash equilibrium by assigning distinct learning rates to the discriminator (4×10^{-4}) and the generator (10^{-4}), an approach we adopted in our PGAN implementation.

2.6 Simulation results

2.6.1 Reconstructions for IC

To evaluate the effectiveness of learned priors in addressing ill-conditioned tomography challenges, we compare the performance of our PGAN against models with no physics assistance and limited physics assistance under constrained imaging conditions of limited angles and low photons, where ill-conditioning is exacerbated. The non-physics-assisted approach, End-to-End training, involves the generative model transforming X-ray projections into imaging objects without incorporating imaging geometry or Poisson statistics. In contrast, the limited physics-assisted strategy utilizes FBP to inform the generative model of the imaging geometry alone. PGAN, however, integrates both the imaging geometry and Poisson statistics through maximum-likelihood estimates as conditional inputs.

For fair comparisons, the generative model and training parameters remain consistent across all methods. Selected IC reconstruction examples under these conditions are depicted in Fig. 2.4, with simulations based on 1800 training and 200 test sets, each comprising $16 \times 16 \times 8$ voxels, under two photon scenarios (256 and 800 photons per ray). Each row in the figure corresponds to a different reconstruction method, with odd columns showing reconstructions at various photon counts and even columns highlighting the absolute differences from the ground truth.

As photon counts increase, generative model-based methods show improved IC reconstruction quality, though improvements are marginal for MLE and FBP methods. Notably, using MLE as conditional input yields superior reconstruction quality, with the FBP-assisted model slightly lagging and the End-to-End method exhibiting the most discrepancies. This suggests the benefits of incorporating physical priors into generative models, which narrows the gap between reconstructed ICs and their ground truths.

Quantitative analysis, as shown in Fig. 2.5, plots the bit error rate of reconstructed 3D IC test datasets against photon counts per ray. For cases between 320 and 800 photons per ray, simulations were repeated with five independent IC circuit sets, with results indicating that the generative model informed by MLE attains single-error-per-sample accuracy at 400 photons. The FBP-informed model requires approximately 800 photons to achieve similar accuracy, while the End-to-End model demonstrates only modest improvements with increased photon counts. This affirms the superiority of the physics-assisted approach in generative modeling for tomographic reconstruction.

To evaluate the design choice of our generative model, we extended our study in a larger range of photon number per ray. Now, variants of the generative models share the same con-

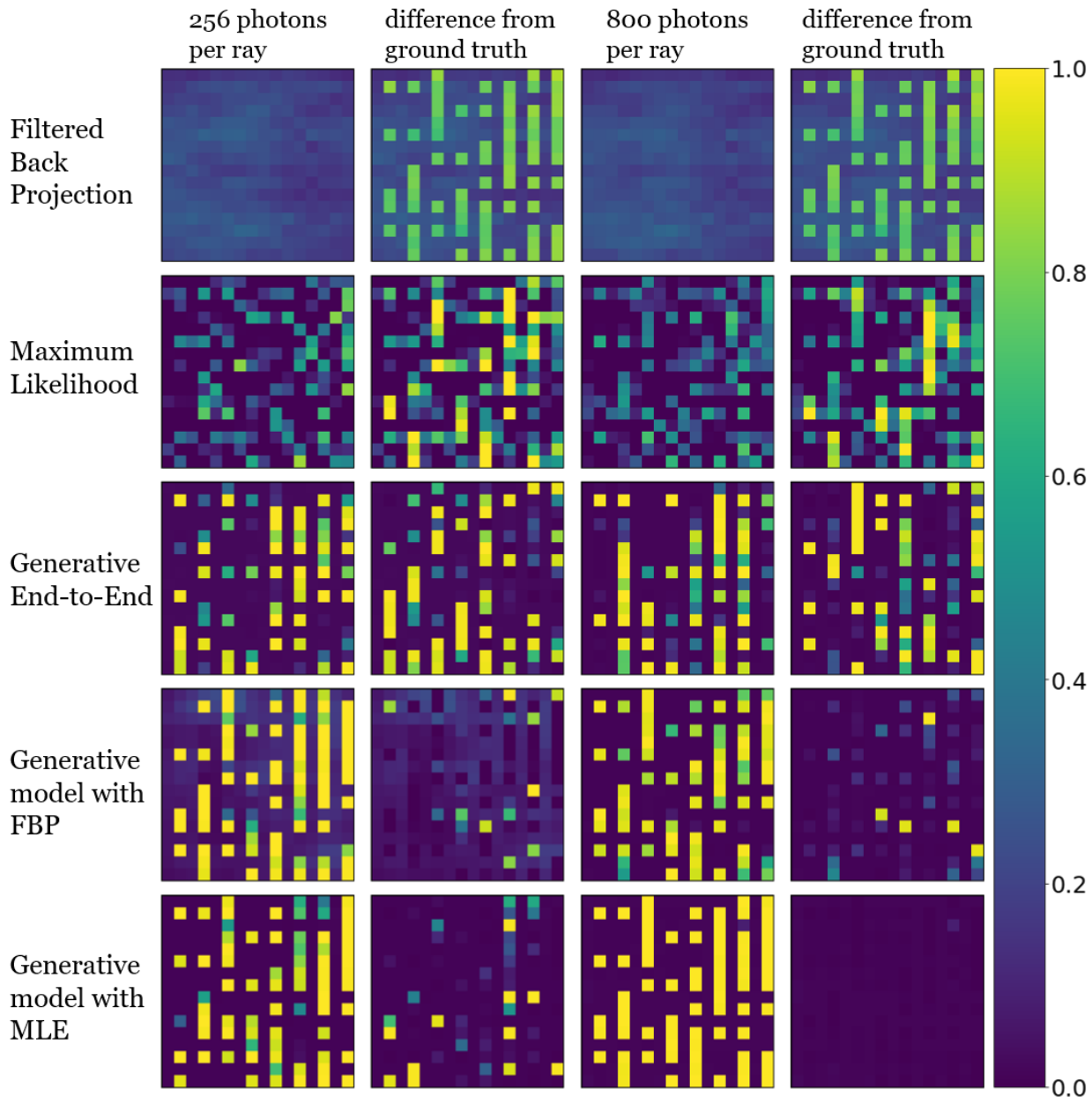


Figure 2.4: Selected examples of IC reconstructions with an angular range of $-30°$ to $22.5°$. The color scale runs from 0 to 1. Each row represents different reconstruction methods, and each column in odd number represents the same location at the given IC distribution with a different photon number per ray, and each column in even number shows the absolute difference between the reconstruction in the previous column to the ground truth circuit.

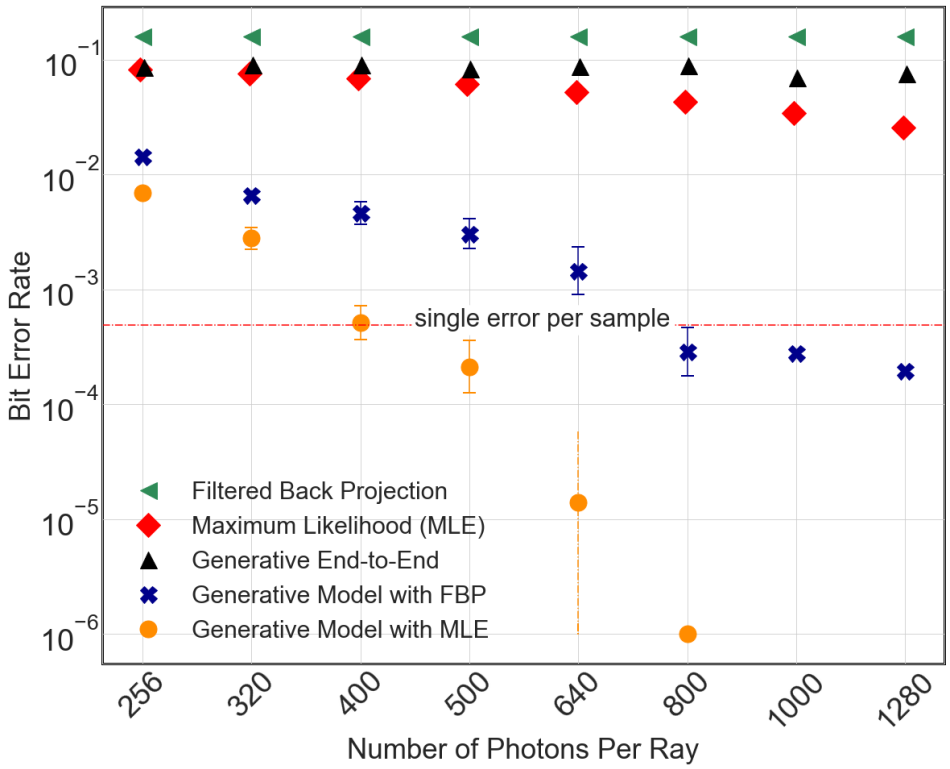


Figure 2.5: Different reconstruction approaches with an angular range of $-30°$ to $22.5°$ for IC objects. Single-error-per-sample in the red dashed line means that the bit error rate is equal to $\frac{1}{16\times16\times8}$.

ditional information (maximum-likelihood reconstructions with the known physical priors) and training parameters, with differences in network architectures.

Figure 2.6 shows a selection of IC reconstructions under limited-angle and low-photon conditions. Each row is a different reconstruction method, with odd-numbered columns illustrating the same IC location at varying photon counts per ray, and even-numbered columns depicting the absolute differences between the reconstruction and the ground truth. Notably, generative reconstructions exhibit improvements over maximum-likelihood reconstructions.

Figure 2.7 provides a quantitative comparison of tomography under limited angles and low-photon conditions, plotting the number of photons per ray (ranging from 100 to 10^4) against the average bit error rate for a reconstructed 3D IC test dataset. Generative models show a critical transition from above to below a single-error-per-sample between 320 and 640 photons per ray. This transition is further analyzed with simulations across five independent synthetic IC circuit sets, where the means and standard errors are reported. At 640 photons per ray, generative models reduce bit error rates by at least two orders of magnitude compared to maximum-likelihood methods, particularly for the model variant with axial attention due to its efficiency in capturing long-range data dependencies. However, generative models using wavelet scattering representation do not demonstrate a performance advantage, likely owing to the convolutional and axial-attention mechanisms already capturing the necessary information from small inputs. Maximum-likelihood reconstructions achieve a single-error-per-sample at approximately 5000 photons per ray, indicating that generative models require

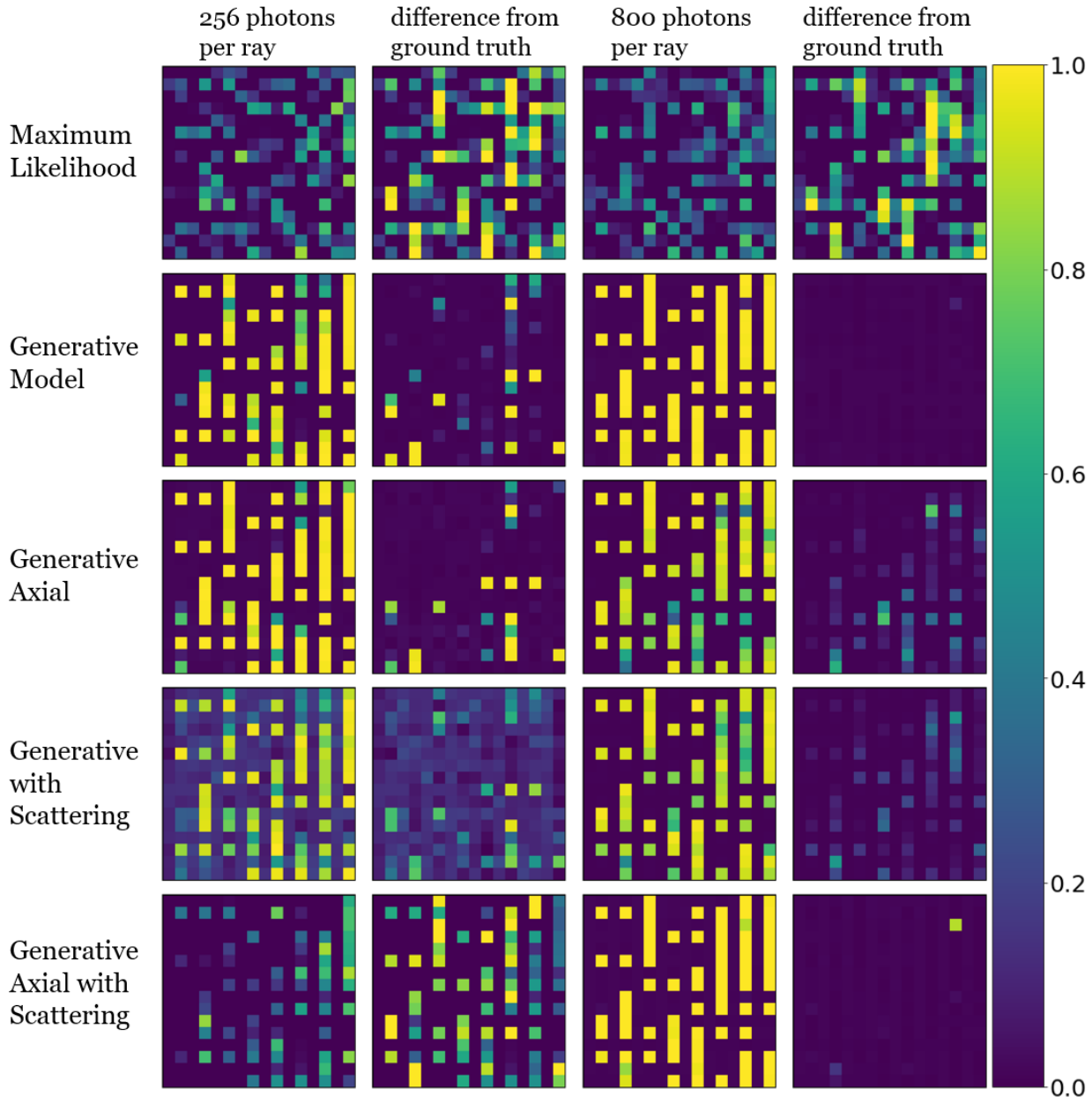


Figure 2.6: Selected examples of IC reconstructions with an angular range of $-30°$ to $22.5°$. The color scale runs from 0 to 1. Each row represents different reconstruction methods, and each column in odd number represents the same location at the given IC distribution with a different photon number per ray, and each column in even number shows the absolute difference between the reconstruction in the previous column to the ground truth circuit.

an order of magnitude fewer photons to reach similar error rates.

We also explores the effectiveness of incorporating a Bayesian prior into the maximum-likelihood algorithm. The selected Bouman-Sauer prior imposes the smoothness of the reconstruction. It is very similar to the Total Variation (TV) prior, particularly since we choose the weights of the neighbors to mimic the absolute value of the gradient. Fig. 2.8

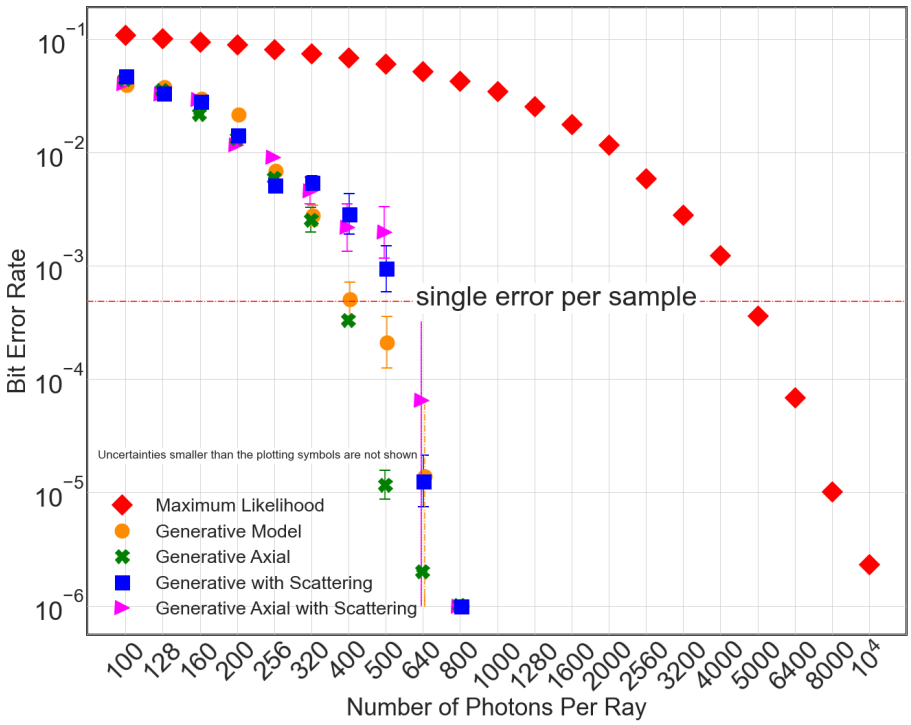


Figure 2.7: Maximum-likelihood vs. generative model reconstructions with an angular range of $-30°$ to $22.5°$ for IC objects.

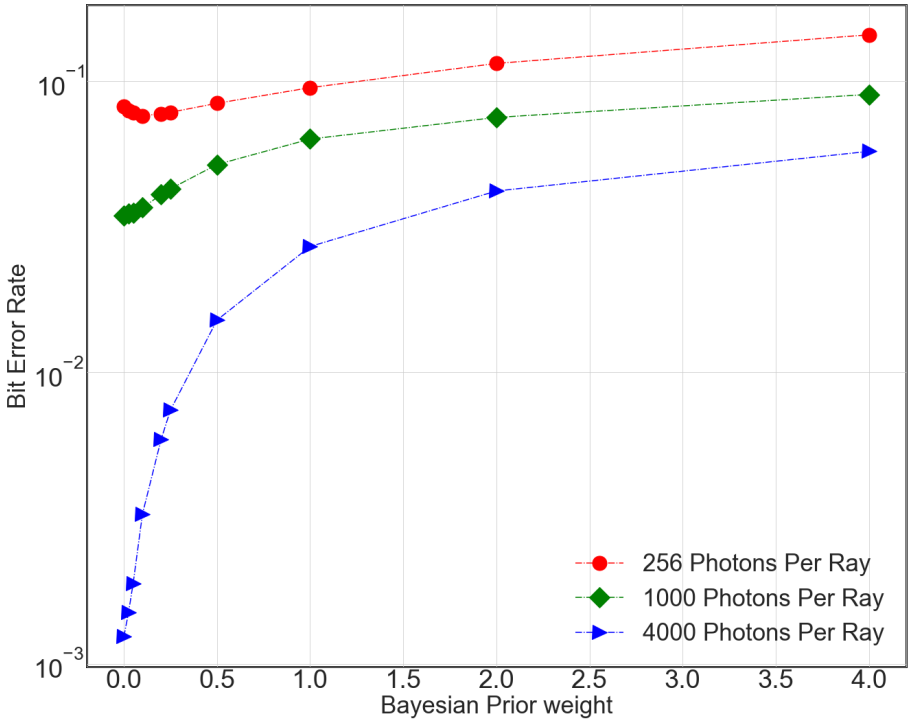


Figure 2.8: Maximum-likelihood reconstructions including the Bouman-Sauer prior with an angular range of $-30°$ to $22.5°$ for IC objects.

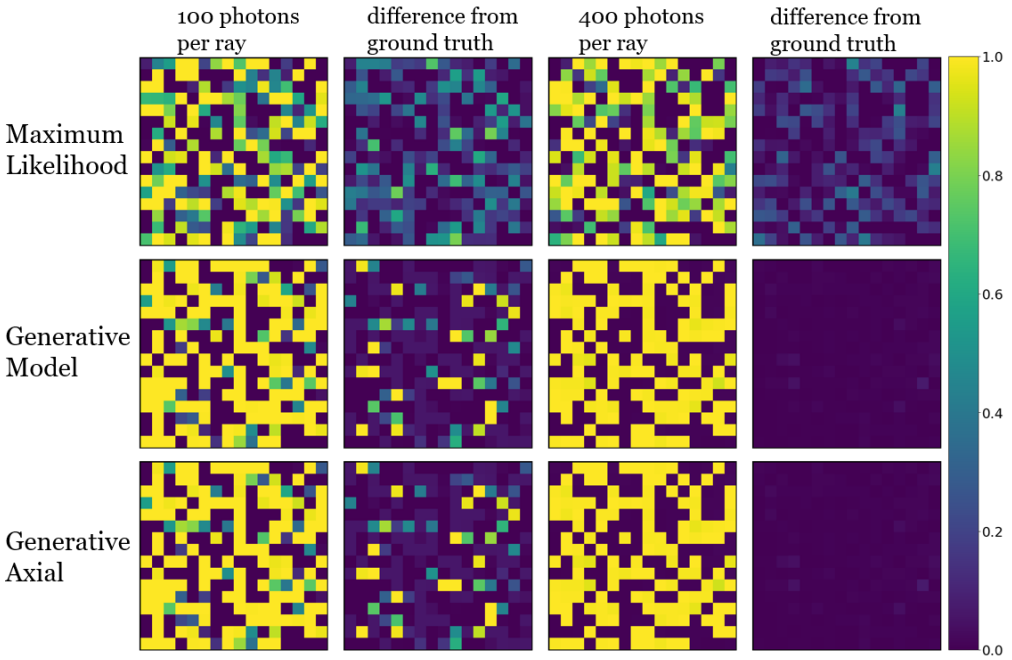
shows the reconstruction quality versus the regularization weight for three imaging conditions with limited angles. The measurements are the 200 test datasets with 3D ICs as the sample, identical to dataset that generates Fig. 2.7. Note that when the weight is 0, the reconstruction is the same as maximum-likelihood estimate without the prior (our baseline in the main text). The Bouman-Sauer prior provides limited improvement for 256 photons per ray case. Quality degradation is obvious for the cases of 1000 and 4000 photons per ray. This outcome suggests that this specific Bayesian prior does not aid our problem.

2.6.2 Reconstructions for independent coin toss object

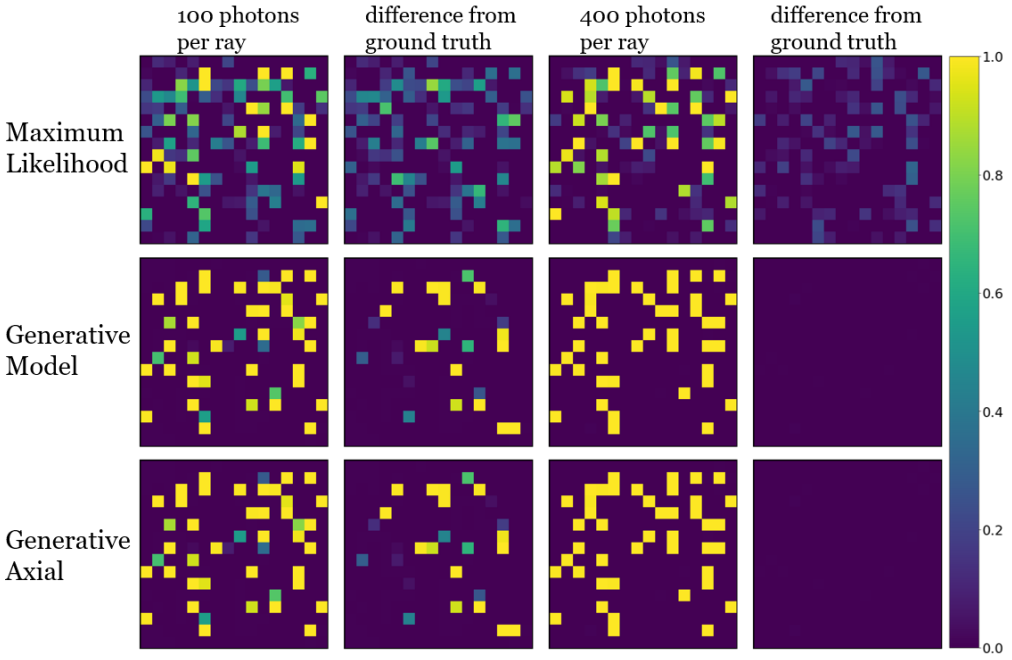
To validate the contribution of the learned prior to the enhancements observed with the generative approach, we extend our quantitative analysis to 3D objects lacking spatial correlation, generated via independent coin tosses for each voxel. With a fair toss probability (copper being 0.5) and an adjusted probability to match the fill fraction of CircuitFaker generated circuits ($p = 0.18521$), the learned prior effectively becomes the probability p for each voxel. Given the absence of spatial correlation, the effectiveness of the learned prior in addressing the inverse problem is anticipated to diminish.

Fig. 2.9 shows the selected examples of independent 3D object reconstruction by limited angle and low-photon tomography. The imaging geometry is the same as before, where the angular range is fixed at $-30°$ to $22.5°$ with $7.5°$ steps. The improvement from the deep generative model is less pronounced than having a circuit object. The assumed prior in our maximum-likelihood approach is now more proper to the reconstruction object with independent voxel. Therefore, the maximum-likelihood estimate improves. The learned prior from the deep generative model behaves similarly to a better classification cut-off for each voxel, shows limited improvements.

Fig. 2.10 shows the quantitative comparison for independent 3D object reconstructions. The x axis is the number of photons per ray in the tomographic projection ranging from 50 to 5000, y axis is the averaged bit error rate of the reconstructed 3D coin toss test dataset. Fig. 2.10(a) is the case with fair coin toss. Compared to the case with spatial correlation (see Fig. 2.7), the required number of photons per ray to achieve single-error-per-sample reduced from 5000 to the range between 320 and 400 for maximum-likelihood estimation. This is attributed to a more proper prior that the maximum-likelihood approach assumed, which leads to better quality reconstructions. The generative models are slightly worse than the maximum-likelihood estimate at lower photon cases. Limited improvements are visible as the generative models need 200 to 256 photons per ray to achieve single-error-per-sample. Fig. 2.10(b) is the case with a biased coin toss that has $p = 0.18521$. Compared with Fig. 2.10(a), the required number of photons per ray for maximum-likelihood estimation to achieve single-error-per-sample is slightly reduced to the range between 200 and 256. With a lower probability of having copper, the 3D objects are now more sparse. Therefore, less attenuation from the copper material leads to effectively more photons captured by the detector pixel, improving the quality of the limited-angle measurements. We also observe a cross-over between maximum-likelihood estimation and deep generative reconstructions in Fig. 2.10 that was not present in the case with imaging object from CircuitFaker. The maximum-likelihood estimation has lower bit error rate in the very low photon regions for independent coin toss objects. This can be explained by our previous interpretation of the

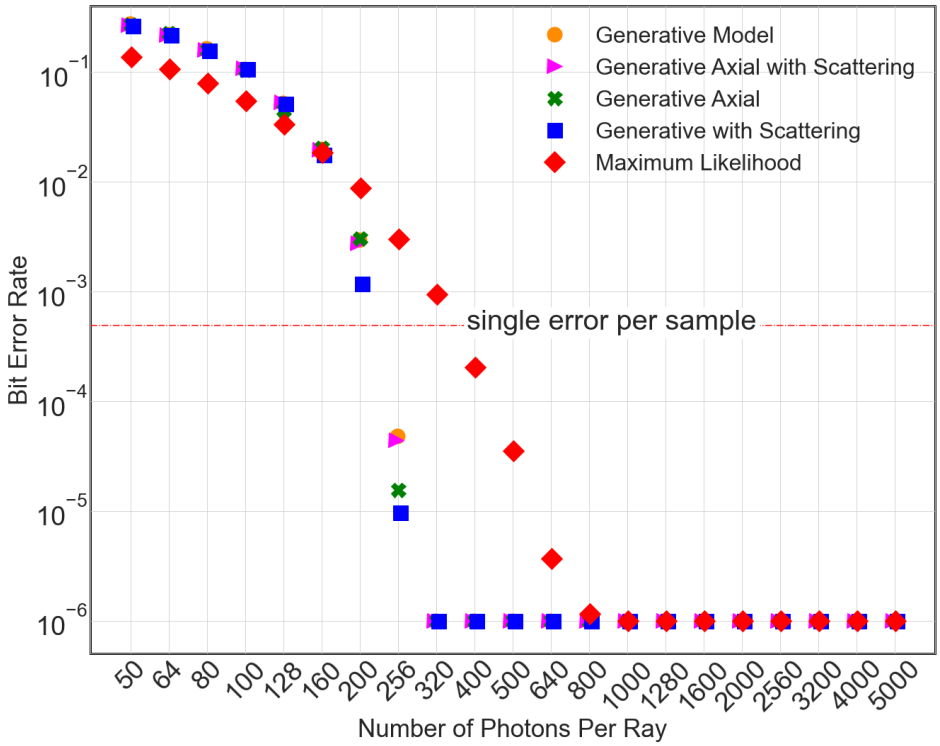


(a) $p = 0.5$ Bernoulli trial (fair coin)

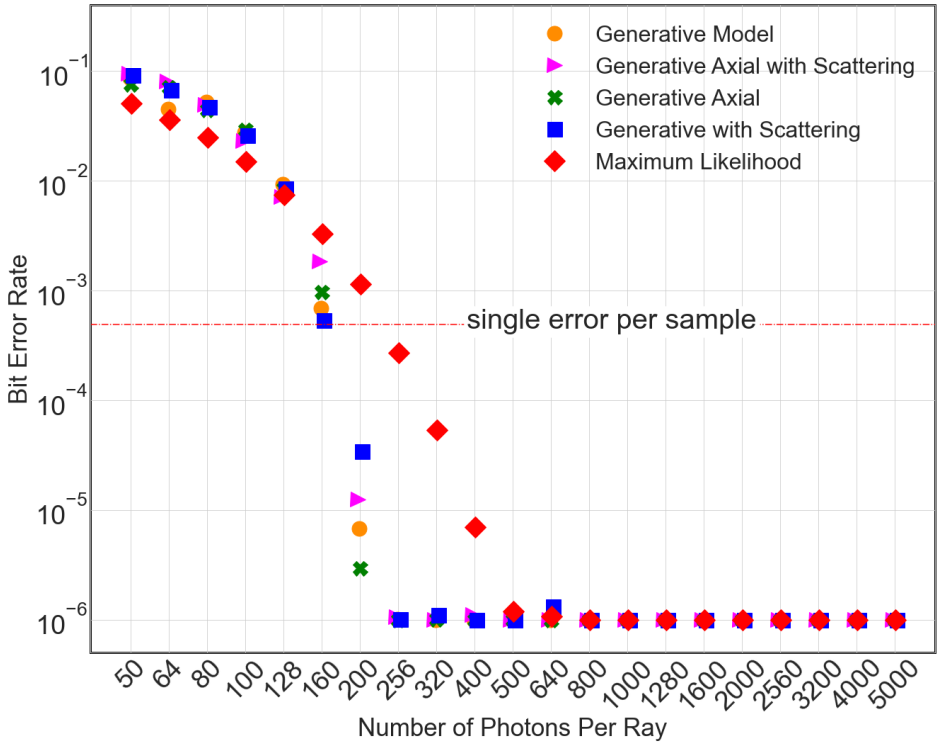


(b) $p = 0.18521$ (unfair coin matching circuit fill factor)

Figure 2.9: Selected examples of independent coin toss an angular range of $-30°$ to $22.5°$. The color scale runs from 0 to 1. Each row represents different reconstruction methods, and each column in odd number represents the same location at the given IC distribution with a different photon number per ray, and each column in even number shows the absolute difference between the reconstruction in the previous column to the ground truth circuit. Reconstructions for Generative Axial at 400 photons per ray resemble the ground truth circuit layer.



(a) $p = 0.5$ Bernoulli trial (fair coin)



(b) $p = 0.18521$ (unfair coin matching circuit fill factor)

Figure 2.10: Results for independent coin toss at every voxel with an angular range of $-30°$ to $22.5°$.

learned prior: since there is no spatial correlation of the object can be learned, a learned classification cut-off may only reduce the quality of a poor reconstruction in the very low photon conditions, especially when all the physical priors have already been considered.

These results confirm that the deep learning approach benefits from the learned prior: when the assumed prior in the iterative algorithm is not well-suited for the reconstruction object (as for the case of circuit reconstruction), the generative models can improve the reconstruction quality. On the other hand, when the prior distribution itself is simple, and the assumed prior matches the distribution (for the independent coin toss object), then the generative models may only provide a marginal improvement over the iterative algorithm.

2.7 Summary and contributions

This study demonstrates the effectiveness of PGAN for X-ray tomography, which integrates maximum-likelihood estimates based on known physics and refines them with a learned prior from a deep generative model. The key contributions of this work include:

1. The development of PGAN, a physics-assisted generative adversarial network for the tomographic reconstruction process. By using a maximum-likelihood estimate derived from an iterative algorithm as input to the generative model, PGAN utilizes both known physics and a learned prior to improve reconstruction quality, especially in limited-angle and low-photon imaging conditions.

2. Demonstrating that further separation of physical priors by using the maximum-likelihood estimate as input, rather than the raw measurements or a direct reconstruction like filtered back-projection, improves the effectiveness of the learned prior from the generative model. This physics-assisted approach reduces the photon requirements to achieve a target reconstruction error compared to methods with no or less incorporation of physical information.

3. Utilizing CircuitFaker, a parametric model developed by Zachary Levine from NIST to generate synthetic 3D circuits emulating real integrated circuit interconnects. CircuitFaker provides a realistic prior distribution for the PGAN model to learn. Using objects generated from CircuitFaker, we show that the learned prior from PGAN enables high-quality reconstructions from limited-angle, low-photon measurements. We thank Zachary Levine for his contribution in creating CircuitFaker and acknowledge his role in its development.

4. Confirming the performance gains are attributable to PGAN's ability to learn the prior distribution of the CircuitFaker model by comparing against reconstructions of objects generated by spatially-independent Bernoulli trials. In this case where the prior is simple and more closely matches the assumptions of the maximum-likelihood estimate, PGAN provides limited improvement, validating that the gains on the CircuitFaker dataset arise from learning the complex prior.

PGAN lowers photon requirements for achieving a specified error rate with limited projection angles. This reduction in photon requirements effectively reduces data acquisition

time during measurement. However, it is important to note that this does not necessarily imply a reduction in the required training data for learning the prior.

The successful application of physics-assisted learned priors in X-ray tomography opens avenues for nanoscale imaging with fewer photons and limited angles. Future work could further explore the impact of PGAN on reducing the amount of training data needed for learning effective priors in X-ray tomography and other imaging modalities. In the following chapter, we will explore one extension of this work that enables data-efficient training for different noise levels encountered during testing.

Chapter 3

Noise-resilient deep learning for integrated circuit tomography

Recent advancements in tomographic reconstruction have incorporated learned regularization priors through machine learning, including unsupervised approaches like sparse coding [123, 124, 125] to regularize the image reconstructions [126, 127], and supervised methods that generate reconstructions directly from measurements [128, 129, 130, 20]. Deep learning, particularly artificial neural networks, has demonstrated success in capturing high-order spatial correlations and noise statistics. Networks based on UNet [131] have significantly improved filtered back-projection (FBP) reconstructions without additional iterations [75]. In Chapter 2, we introduced a physics-assisted generative adversarial network (PGAN) that improves reconstruction quality at low photon budgets compared to alternative methods with less physics assisted approaches [28]. However, while effective, PGAN is specific to the noise level used during training, raising the question of how to handle varying noise levels between training and testing.

Researchers in the machine learning community have investigated this problem, known as out-of-distribution generalization [132, 133, 134]. In computational imaging, studies have shown that learned priors face challenges in noise resilience under out-of-distribution conditions, leading to unstable and low-quality reconstruction [22, 79, 23]. Therefore, when the noise level of the measurements varies, training multiple networks to handle different noise levels for optimal reconstruction quality is necessary, which is a limitation of PGAN.

In this chapter, we introduce a noise-resilient deep-reconstruction algorithm for X-ray tomography of integrated circuits. By incorporating maximum *a posteriori* (MAP) reconstructions with Gaussian noise and sparsity-promoting priors into the neural network input, we mitigate input distribution shifts from varying noise levels, improving the noise resilience of the learned prior. Unlike PGAN, our approach only requires training one network at a particular noise level and can generalize to different noise levels at test time, making it particularly advantageous in scenarios like circuit imaging, where obtaining training datasets with varied noise levels is time-consuming and difficult.

We demonstrate the effectiveness of our noise-resilient deep learning, which utilizes a Total Variation regularizer on MAP estimates, in improving noise resilience through simulations and experiments. Our method achieves high fidelity reconstructions with fewer photons compared to traditional FBP+UNet methods for integrated circuit imaging, with up to an

8× reduction in required photons in simulations and a 2.5× reduction in experiments. These findings highlight the potential of our algorithm in enhancing the efficiency and quality of X-ray tomography for integrated circuit imaging while maintaining data efficiency by eliminating the need for diverse noise-level training data.

3.1 Deep-reconstruction network

Supervised learning uses training data to capture the underlying patterns of a given problem. In the context of tomography, this involves learning from pairs of physical measurements g^* and corresponding ground truths f_0, drawn from a joint distribution $P(g^*, f_0)$. The goal is to approximate the conditional probability $P(f_0 \mid g^*)$, essentially learning to predict f_0 from g^* [128, 129, 130, 20]. This task transforms into an optimization problem aiming to identify optimal network parameters (weights) $\mathbf{w}$ that map g^* to f_0. The process, encapsulated by

$$\widehat{\mathbf{w}}_g = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}_{(g^*, f_0)} \left[\mathscr{L} \left\{ G_{\mathbf{w}}(g^*),\, f_0 \right\} \right] \approx \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{n_{\text{train}}} \mathscr{L} \left\{ G_{\mathbf{w}}(g_i^*),\, f_{0i} \right\}, \tag{3.1}$$

where $G_{\mathbf{w}}$ denotes the neural network, n_{train} the number of training samples, and $\mathscr{L}$ the loss function, is known as end-to-end training.

Physics-assisted training introduces an additional step by preprocessing the input data with a conventional reconstruction algorithm before feeding it into the neural network [75, 24, 40, 28]. This method effectively lightens the network's learning, expressed as

$$\widehat{\mathbf{w}}_f = \underset{\mathbf{w}}{\operatorname{argmin}}\, \mathbb{E}_{(\hat{f}, f_0)} \left[\mathscr{L} \left\{ G_{\mathbf{w}}(\hat{f}),\, f_0 \right\} \right] \approx \underset{\mathbf{w}}{\operatorname{argmin}} \sum_{i=1}^{n_{\text{train}}} \mathscr{L} \left\{ G_{\mathbf{w}}(\hat{f}_i),\, f_{0i} \right\}, \tag{3.2}$$

with $\hat{f}$ representing the conventional algorithm's object estimate.

A widely adopted approach is combining FB) with a convolutional neural network such as UNet. It utilizes FBP-generated approximations to incorporate the imaging system's physical model directly. Post-training, the UNet is used in artifact removal and image structure preservation within FBP reconstructions, sidestepping the need for further iterations and regularization. This approach has been adopted for its computational efficiency, low latency, and superior reconstruction quality compared to conventional methods like MAP with general-purpose priors [75, 135, 136].

3.2 Noise resilience of deep-reconstruction networks

A major concern in employing supervised learning for tomography is the generalization problem, specifically, the adaptability of the learned prior to new, unseen test data. If optimally trained, such priors can yield high-quality reconstructions for test data matching the training data's distribution. However, in applications, discrepancies often exist between the distributions of training and test data. Despite knowing the imaging object class and maintaining consistent imaging geometry, noise characteristics within the data may fluctuate over time.

Prior works on supervised learning typically assume the noise statistics in training and test datasets are similar. For instance, if Poisson noise corrupts measurements g^*, the evaluation of a learned prior's reconstruction quality is based on test data with identical photon count N_0 per ray as in training. However, when test data has a different photon count N_1 per ray, a shift in noise distribution between training and testing phases arises, potentially compromising the learned prior's reconstruction accuracy without training samples at N_1 photons. The issue lies in the potential unreliability of the estimated conditional probability $P\left(f_0 \,|\, g^*\right)$ if $P\left(g^*_{N_0}, f_0\right) \not\approx P\left(g^*_{N_1}, f_0\right)$, where N_0 and N_1 represent varying photon statistics in measurements.

To improve this noise resilience, one approach might involve using a range of distributions $P\left(g^*_{N_i}, f_0\right), N_i \in \{0, 1, 2, ...\}$, thus creating a network adaptable across noise statistics. Alternatively, constructing a series of $P\left(f_0 \,|\, g^*_{N_i}\right), N_i \in \{0, 1, 2, ...\}$, could yield a collection of networks, each tailored to a specific noise level. However, these strategies demand extensive training data across various noise levels, a daunting task for tomography systems characterized by lengthy acquisition times. Our method circumvents the need for additional noisy data collection by training with a singular joint distribution. Employing a Gaussian noise prior and a sparsity-promoting prior for reconstructions, we generate noise-resilient $\hat{f}_{\text{MAP}}$ through maximum *a posteriori* estimation. Subsequently, we approximate $P\left(f_0 \,|\, \hat{f}_{\text{MAP}}\right)$ by sampling the training distribution $P\left(\hat{f}_{\text{MAP}}, f_0\right)$ at a fixed photon statistic level. This approach allows us to assess the learned prior's noise resilience using test data across a spectrum of photon statistics.

3.3 Evaluation methods

3.3.1 Description of X-ray tomography experiment

Our study employs the Zeiss Xradia 620 Versa X-ray imaging system located at MIT.nano. This system generates X-rays using a tungsten target, operating at a tube voltage of 80.0 kV and a power of 10.0 W. We utilize a 3D printed sample designed by CircuitFaker [28], which simulates basic circuit-like structures in three dimensions. The sample is positioned 230 mm from the source and undergoes a full rotation from 0° to 360° in 1600 steps, facilitated by a precision rotation stage. A charge-coupled device (CCD) detector captures the projections, situated 572 mm from the source. Given a Fresnel number around 100, the projection-based forward model is deemed suitable for our experiments. Both simulation and experimental setups mirror this imaging geometry, employing a cone angle of 3° (with a maximum divergent angle of 1.5°) and accommodating a maximum tolerance angle of approximately 2.4° for the sample. This configuration validates the parallel-beam (Radon) approximation for our forward model. For the simulations, we generate Poisson-distributed noise to replicate measurement statistics. Experimentally, the exposure time ranges from 35 ms to 9 s per projection, allowing for a controlled variation in photon counts.

3.3.2 Reconstruction algorithms for comparison

We compare three traditional tomographic reconstruction algorithms, FBP, MLE, and MAP with a total variation prior (MAP/TV), against their enhanced counterparts that incorporate a UNet architecture for further refinement, as FBP+UNet, MLE+UNet, and MAP+UNet. Our primary contribution is the MAP+UNet approach, designed to improve noise resilience by refining MAP reconstructions through the approximation of $P\left(f_0 \,|\, \hat{f}_{\text{MAP}}\right)$. In contrast, MLE+UNet and FBP+UNet improve upon their foundational algorithms by approximating $P\left(f_0 \,|\, \hat{f}_{\text{MLE}}\right)$ and $P\left(f_0 \,|\, \hat{f}_{\text{FBP}}\right)$ respectively. The algorithms' acronyms are detailed in Table 3.1 for clarity. To ensure a balanced evaluation, all learning-based models utilize identical optimization parameters and share the same UNet framework, differing only in the initial reconstruction inputs. A consistent regularization parameter ($\beta = 2$) applies across MAP reconstructions at varying noise levels. This setup is illustrated in a conceptual diagram (Fig. 3.1). For simulations, network weights are initially randomized and subsequently refined using 10,000 noise-free training samples. Experimental model weights initialized with values derived from simulation data, then fine-tuned with 120 experimental samples, a method known as transfer learning. This approach effectively minimizes the volume of experimental data necessary for training [137].

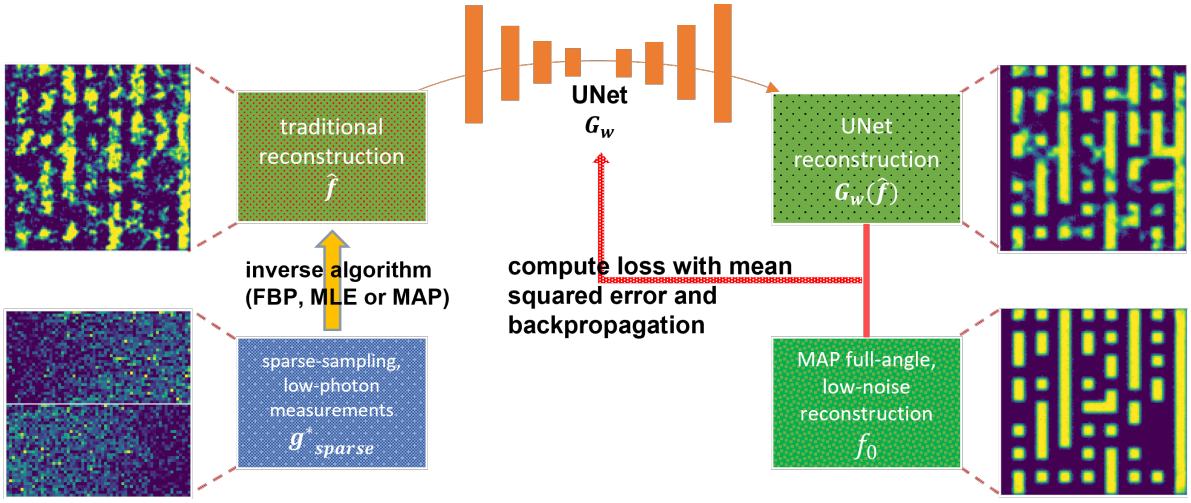


Figure 3.1: A conceptual diagram for the learning-based algorithms. An inverse algorithm first produces the traditional reconstruction $\hat{f}$ from the sparsely-sampled and low photon measurements. Then the UNet takes $\hat{f}$ and outputs reconstruction $G_{\mathbf{w}}(\hat{f})$.

3.3.3 Network architecture

Fig. 3.2 shows the network architecture for the learning-based algorithms. Four downsampling blocks and four up-sampling blocks were used in the UNet-like architecture. The spatial dimension of the feature map is reduced or up-sampled by 2 per block. The downsampling is achieved by stride convolution and the up-sampling is by transposed convolution. The initial input is $128 \times 128 \times 1$, the same as the output dimension. The latent vector in the

Table 3.1: Acronyms for the reconstruction algorithms.

Abbreviation	Definition
FBP	filtered back projection
MLE	maximum likelihood estimate
MAP/TV	maximum *a posteriori* estimate with Total Variation
FBP+UNet	improved FBP, MLE, or MAP/TV reconstruction using the learned prior from UNet, shown in Fig. 3.1
MLE+UNet	
MAP+UNet	

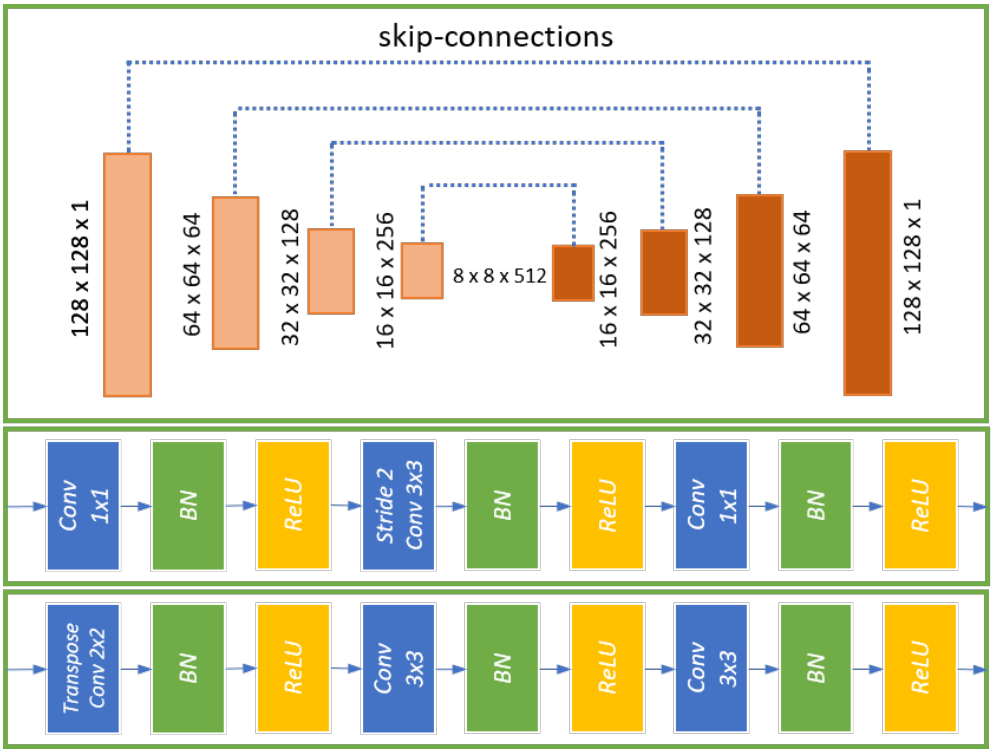


Figure 3.2: UNet architecture for the learning-based algorithms. The top box shows the overall design of the network, where the light orange modules are downsampling blocks and the dark orange modules are up-sampling blocks. Blue dotted lines represent the skip connections. The middle box shows the design of the downsampling blocks. The bottom box shows the design of up-sampling blocks. BN is batch normalization.

center has the dimension $8 \times 8 \times 512$, where the channel size is in the last dimension. ReLU is used as the activation function in the blocks. The skip connections concatenate features from downsampling blocks to up-sampling blocks. The script that generates the architecture can be found on our GitHub page [138].

3.3.4 Quality metrics and their acceptability thresholds

Pearson correlation coefficient

The Pearson correlation coefficient r is defined as

$$r_{f_0,\hat{f}} = \frac{\mathrm{cov}(f_0, \hat{f})}{\sigma_{f_0}\, \sigma_{\hat{f}}}, \tag{3.3}$$

between ground truth reconstruction f_0 and reconstruction $\hat{f}$ from a particular algorithm, where cov is the covariance and σ is the standard deviation. This first metric is introduced to evaluate the perceptual quality of the reconstruction, and the acceptable quality threshold is $1 - r \leq 10^{-1}$, indicating a strong linear relationship between the reconstructions [139]. Its shortcoming is that it is a pixel-by-pixel correlation that is sensitive to misregistration and image distortion, and is not highly sensitive to the connectivity or topology of the image [140].

Mallat Scattering Transformation

The normalized L^2 distance of the logarithm of the Mallat Scattering Transform (MST) space [105, 141] is defined as

$$\varphi_{f_0,\hat{f}} = \frac{\|\Phi(f_0) - \Phi(\hat{f})\|^2}{\|\Phi(f_0)\|\, \|\Phi(\hat{f})\|}, \tag{3.4}$$

where Φ is the logarithm of the MST operator. This second metric is introduced to evaluate multi-scale correlations and topology (connectivity) of the reconstructions. It is also insensitive to misregistration and image distortion. The acceptable quality threshold is $\varphi \leq n/M^2 = 3 \times 10^{-3}$, where the reconstruction is of size $M^2 = 128 \times 128$ and $n = 50$ is the average number of circuit elements.

MST can be viewed as a Convolutional Neural Network (CNN) with predetermined weights. The filters are designed so that the CNN can span an exponentially large range in scale with a kernel of constant size. Following [142], we define the logarithm of MST of an input image f as

$$\Phi(f) = \left(\log \Phi_J^{(0)}(f), \log \Phi_J^{(1)}(f), \log \Phi_J^{(2)}(f)\right) \tag{3.5}$$

and

$$\begin{aligned}
&\Phi_J^{(0)}(f) = \phi_J \circledast f \\
&\Phi_J^{(1)}(f) = \phi_J \circledast \left|\, \psi_{\lambda_1} \circledast f \,\right|, && \lambda_1 \in \Lambda_1, \\
&\Phi_J^{(2)}(f) = \phi_J \circledast \left|\, \psi_{\lambda_2} \circledast \left|\, \psi_{\lambda_1} \circledast f \,\right| \right|, && \lambda_1 \in \Lambda_1, \lambda_2 \in \Lambda_2.
\end{aligned} \tag{3.6}$$

Here, $\circledast$ denotes the convolution in 2D space, ϕ_J is a low-pass filter, $\{\psi_{\lambda_i}\}$ is a family of band-pass filters, $i = 1, 2$. Morlet filters [143] are used in our computation. There are three parameters that determine how the MST is taken: M^2 is the number of pixels in the image, J is the $\log_2$ of the scattering scale and L is the number of angles used in the transform. For this work, values of $M = 128$, $J = 4$, and $L = 8$ were used.

MST has been shown by Mallat [105] to be Lipschitz continuous to diffeomorphic deformation and invariant under translation. Therefore, MST is insensitive to misregistration and deformation because they are both small diffeomorphic deformations. It also induces a high sensitivity to topology, since the topology is invariant to diffeomorphic deformation. Furthermore, taking the logarithm of the transformation flattens the extracted features to a low-dimensional complex linear subspace where the topology of the features is exposed. The practical outcome is that the extracted features in MST space will form a high-precision cluster. The precision should be the fractional dimension of the space, n/M^2, where n is the number of the features.

3.4 Simulation results

Our simulation-based investigation showcases the efficacy of our noise-resilient methodology in addressing ill-conditioned tomography challenges. We employed a full-angle, sparse sampling setup with 32 out of 1600 angular views evenly distributed, under conditions of low-photon tomography where the absence of regularization prominently exacerbates ill-conditioning. The visual outcomes, depicted in Fig. 3.3, present 2D reconstructions at varying photon counts per ray for distinct algorithms. Each row in the figure shows the performance of different reconstruction algorithms, while each column illustrates the impact of varying photon counts per ray on reconstructions. Notably, the MAP+UNet approach outperforms both MLE+UNet and FBP+UNet, especially at reduced photon counts per ray, demonstrating superior reconstruction quality.

A quantitative analysis, illustrated in Fig. 3.4, aggregates the mean values and standard errors across two metrics over 1000 test instances from CircuitFaker. Observations from both metrics converge on similar trends, prompting a focused discussion on the MST metric. With a benchmark threshold of 3×10^{-3}, the MAP+UNet strategy meets the criteria for photon fluxes exceeding 80 photons per ray, establishing itself as the most effective method. Following in effectiveness, the MLE+UNet approach meets the threshold for photon counts of 128 per ray and above. Among the evaluated learning-based algorithms, FBP+UNet displays the lowest noise resilience, tolerating noisy conditions down to 640 photons per ray. Consequently, the MAP+UNet approach facilitates an 8× reduction in photon usage compared to FBP+UNet in our simulations.

Furthermore, under low photon flux conditions, the sparsity-promoting MAP+UNet outperforms the MLE+UNet based on maximum likelihood. However, this performance gap narrows with the increase in photon flux. Importantly, integrating learning via a UNet with traditional reconstruction algorithms reduces the photon flux necessities, suggesting the value of the learned priors.

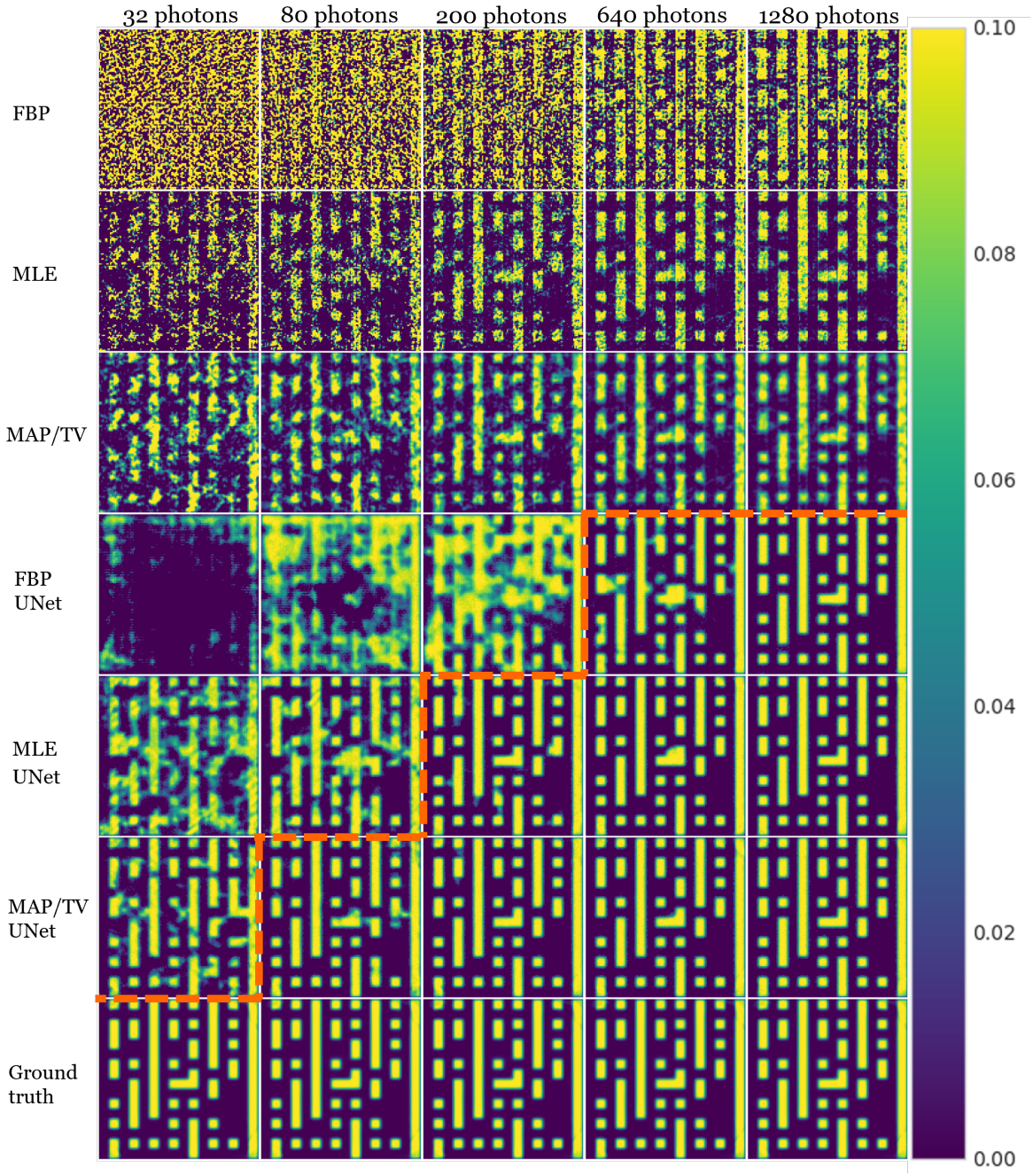


Figure 3.3: Selected 2D reconstructions (in 128×128) for different algorithms using simulated data. Each row represents a reconstruction algorithm, and each column represents an intensity of the photon rays. The ground truth is repeated in the last row. The dotted orange line is the boundary between acceptable and unacceptable performance as determined by the MST metric.

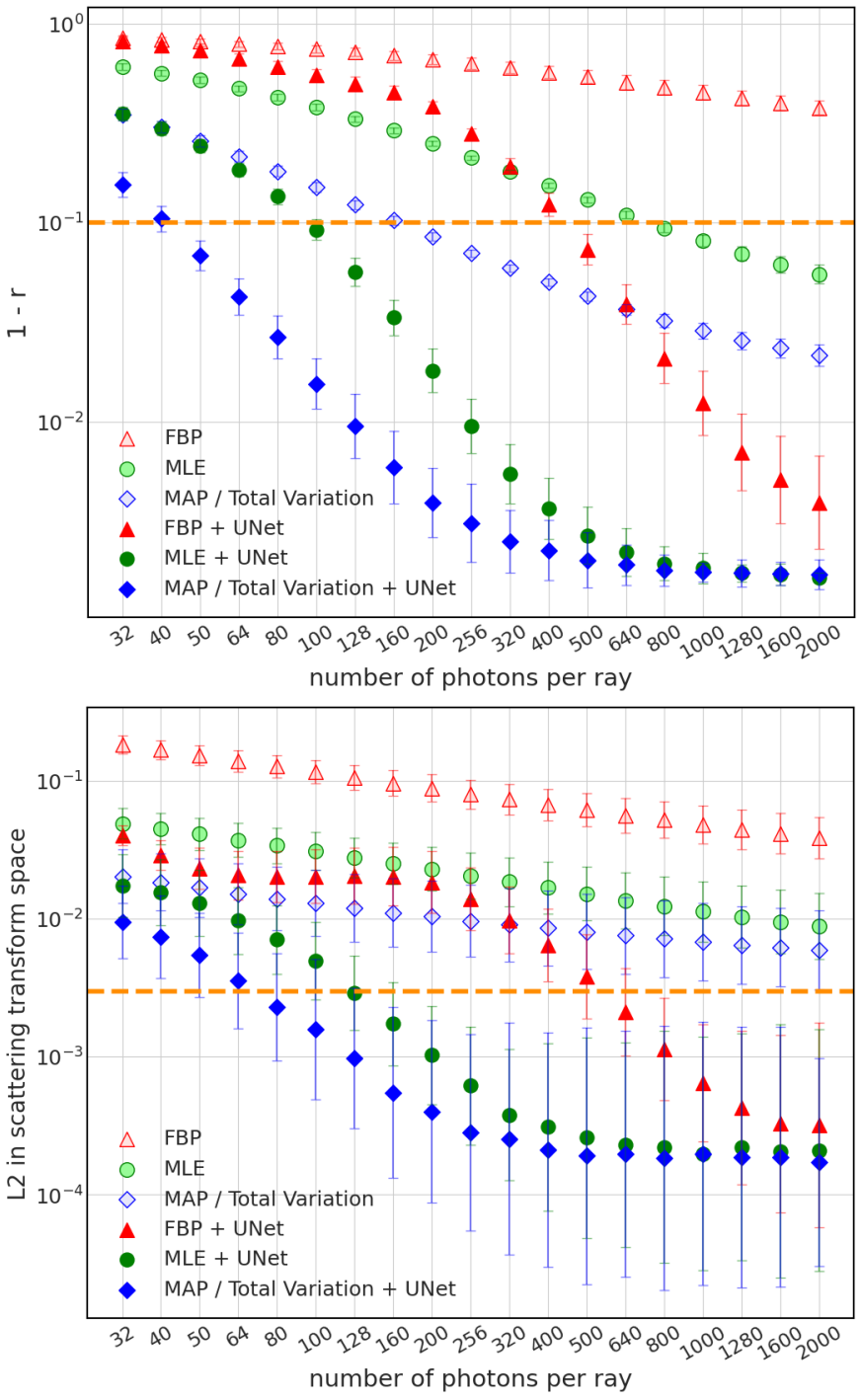


Figure 3.4: Quantitative comparison between different reconstruction algorithms for tomographic simulations under different photon counts per ray. The x axis is the number of photons per ray, and the y axis on the left figure is $1 - r$ where r is the Pearson correlation coefficient. The y axis on the right is the L^2 distance in MST. The error bars are standard deviations in the log scale of 1000 test instances. The dotted orange line shows the thresholds of acceptable performance.

3.5 Experimental results

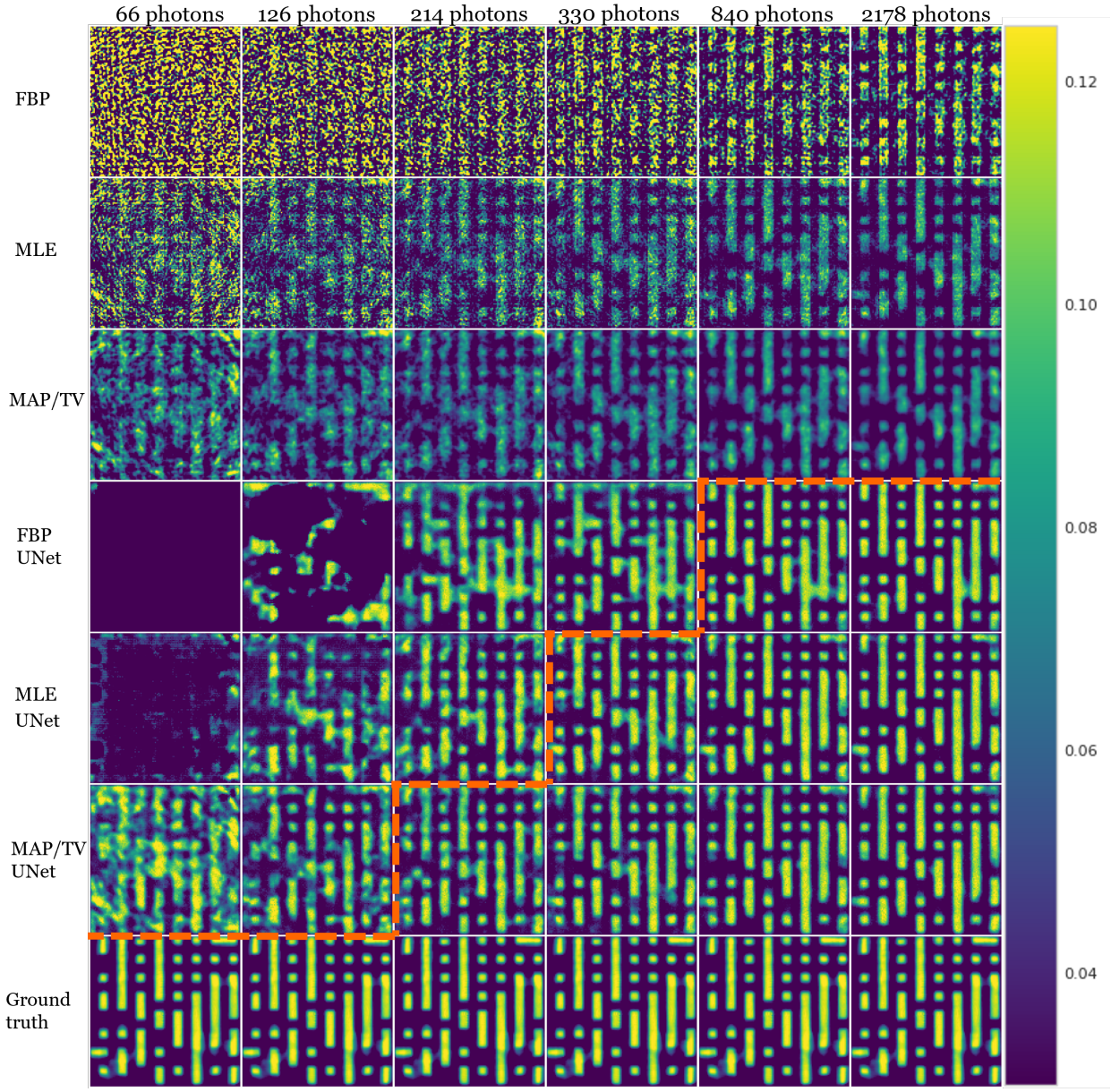


Figure 3.5: Selected 2D reconstructions (in 128×128) for different algorithms using experimental data. Each row represents a reconstruction algorithm. Each column represents an intensity of the photon rays. The dotted orange line is the boundary between acceptable and unacceptable performance as determined by the MST metric.

We extended our investigation of the noise-resilient approach to experimental datasets under identical imaging conditions as those simulated. 2D reconstructions from experimental projection data at varying photon counts per ray are depicted in Fig. 3.5, showcasing the performance of different algorithms.

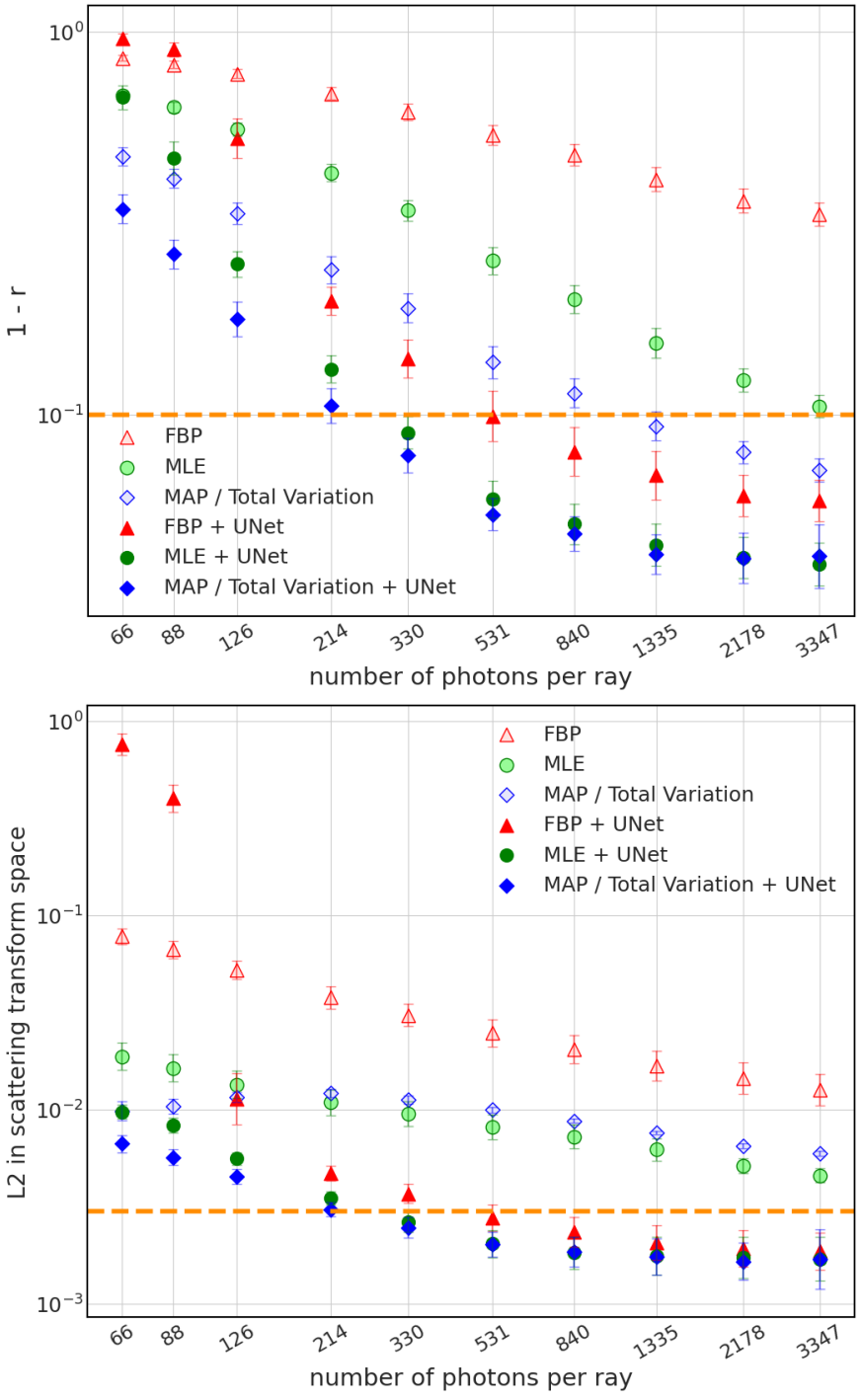


Figure 3.6: Quantitative comparison between different reconstruction algorithms for experimental data of 40 instances under different photon counts per ray. Symbols and error bars as in Fig. 3.4. The dotted orange lines show the thresholds of acceptable performance.

Quantitative assessments of these algorithms, derived from experimental data across a range of photon fluxes, are presented in Fig. 3.6. These evaluations, based on two quality metrics, affirm the simulation findings: superior input reconstructions yield equally or more accurate UNet reconstructions, with parity more probable at higher photon counts. At diminished photon counts, MAP+UNet emerges as the most noise-resilient among the learning-based algorithms. Adopting the same quality benchmark from the simulation ($\varphi \leq 3 \times 10^{-3}$ for the MST metric), traditional algorithms fail to reach acceptable quality under sparse sampling and low-photon conditions. Conversely, all learning-based algorithms achieve satisfactory outcomes. Table 3.2 consolidates the performance thresholds for all learning-based algorithms, both experimentally and in simulation.

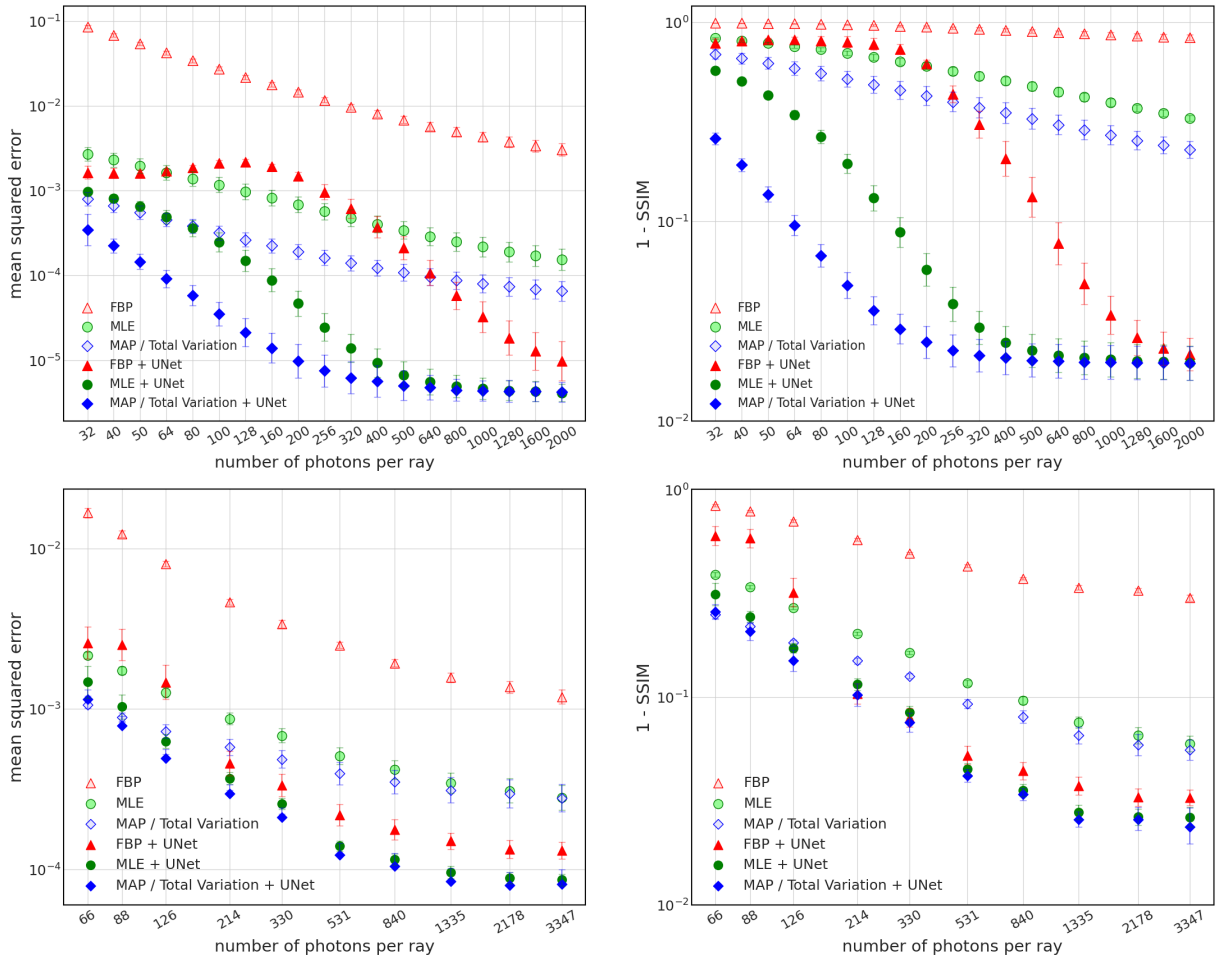


Figure 3.7: Quantitative comparison between different reconstruction algorithms with mean squared error (MSE) and structural similarity index measure (SSIM) metrics. The top two figures are for simulated data, and the bottom two are for experimental results.

In Fig. 3.7, we present our quantitative comparison of various reconstruction algorithms with mean squared error (MSE) and structural similarity index measure (SSIM) metrics. We find that our main conclusions do not depend on which metric is chosen.

In essence, MAP+UNet enables a reduction in photon usage by 2.5× in experimental settings compared to FBP+UNet, a contrast to the 8× reduction observed in simulations.

Table 3.2: Thresholds for acceptable performance based on the MST metric.

Method	**Simulation** (photons/ray)	**Experiment** (photons/ray)
FBP+UNet	640	531
MLE+UNet (+Gaussian)	128	330
MAP+UNet (+Gaussian+sparsity)	80	214

3.6 Discussion

In this investigation, Total Variation (TV) was adopted as the sparsity-promoting prior for MAP and MAP+UNet algorithms. Although TV effectively reconstructs simple circuit structures, its efficacy for complex structures in varied imaging tasks may be constrained. Exploring alternative priors, such as wavelet-based and Laplacian, could improve our method's versatility across diverse applications [144, 145, 146]. Additionally, the regularization parameter for TV was not optimized for different noise conditions, indicating potential for improvement by adjusting this parameter to match the noise level.

The alignment between simulation and experimental results underscores the method's validity, albeit with noted performance threshold discrepancies. Such variations likely stem from real-world complexities like additional noise sources, system imperfections, and model simplifications that simulations cannot fully replicate.

3.7 Summary and contributions

We present a noise-resilient deep learning approach for X-ray tomography of integrated circuits, validated through simulations and experiments. Our method incorporates MAP reconstructions as training inputs to obtain a learned prior that withstands noise variations without requiring diverse noise-level training data. This data-efficient approach is valuable when acquiring varied datasets is impractical due to long acquisition times. Furthermore, it is particularly useful in real-time dynamic imaging scenarios. Our main contributions are:

1. Demonstrating that using MAP reconstructions with Gaussian noise and Total Variation priors as inputs to a convolutional neural network improves noise resilience compared to FBP or MLE reconstructions.

2. Thoroughly examining and addressing generalization of deep learning reconstruction algorithms to out-of-distribution test data with different noise levels than training data. Achieving noise resilience during the testing phase without requiring additional training data at various noise levels, crucial for applications like integrated circuit imaging with limited dataset sizes.

3. Quantitatively evaluating noise resilience using Pearson correlation and Mallat Scattering Transformation metrics, establishing performance thresholds. MAP+UNet reduces photons per ray by 8× in simulations and 2.5× in experiments compared to FBP+UNet.

Our approach advances X-ray tomography by enabling high-quality reconstructions with reduced photon requirements, improved noise robustness, and data efficiency, enhancing deep learning for X-ray imaging in real-world scenarios with different training and test noise distributions.

It is worth noting that while the physics-assisted generative adversarial network introduced in Chapter 2 also reduces photon requirements, it requires training a new network for each noise level. In contrast, the noise-resilient deep learning approach presented in this chapter only requires training one network at a particular noise level, and it can generalize to different noise levels at test time without the need for retraining. This feature makes our noise-resilient approach more practical and efficient in scenarios where the noise level may vary during the imaging process or across different datasets, reducing computational cost and increasing flexibility in real-world applications.

Chapter 4

Data-efficient neural modeling for 3D Printing

Two-color projection micro-stereolithography (PμSL) is an additive manufacturing technique that enables rapid, continuous printing of 3D objects [147, 148, 149]. This approach utilizes two different wavelengths of light to independently control photoinitiation and photoinhibition of polymerization in a photocurable resin. By carefully selecting the photoinitiator and photoinhibitor with complementary absorption spectra, the two processes can be controlled independently, allowing for continuous printing, high precision, and the ability to use viscous resins. The technique finds potential applications in the development of optical diffractive neural networks (D2NNs), which necessitates the manufacturing of sophisticated three-dimensional photonic structures capable of optical computation [150, 151, 152, 153].

Despite its potential, achieving high-precision 3D printing with PμSL remains challenging due to the limitations of current inspection and modeling approaches. The photochemical processes that are present during photopolymerization are extremely complex, and that an understanding of these processes is importance if a practical model is to be developed [154, 155, 156]. Conventional inspection techniques, such as profilometry, atomic force microscopy, or X-ray tomography, are typically performed offline after printing, washing, and cleaning steps, hindering rapid iteration and feedback to the printing process. These limitations make it difficult to achieve the level of precision required for advanced applications like optical neural networks.

In this chapter, we introduce a data-efficient approach that employs a neural network to model and optimize the diffusion processes within the two-color PμSL manufacturing. Our neural model, derived from ODEs with a data-driven correction term modeling the non-local diffusion effects, offers a more accurate representation of the chemical conversion dynamics during the printing process. Notably, the ODE component of the model is pre-determined by conventional Fourier-transform infrared spectroscopy (FTIR) measurements [157, 158], contributing to the data efficiency of the neural network. By integrating coherent diffractive phase imaging into the printing system, we can collect the 2D dynamics of the conversion process in real-time. Additionally, we employ conventional microscope imaging after the printing process to capture the final printed structure. We train the neural model to predict the final printed structure given the printing parameters, effectively modeling the non-local diffusion from a series of 2D measurements.

Through simulations and experimental validation, we demonstrate the advantages of our neural model over conventional models in achieving better precision in the fabrication of optical diffractive elements using two-color PμSL. The combination of real-time phase imaging and post-printing microscope imaging allows our model to capture the complex diffusion dynamics, resulting in improved precision in the manufacturing process. This synergistic approach not only advances the state of the art in high-precision 3D printing but also holds the potential to drive breakthroughs in the development of next-generation optical neural networks and computational architectures.

4.1 Projection micro-stereolithography for 3D printing

Projection micro-stereolithography (PμSL) is an advanced 3D printing technique based on area projection triggered photopolymerization, capable of fabricating complex 3D architectures covering multiple scales with precision up to 0.6μm [159]. PμSL stands out among various 3D printing technologies for its ability to achieve high precision over relatively wide areas. This technique utilizes a digital micromirror device (DMD) to modulate UV light and project 2D patterns onto a photosensitive resin, selectively polymerizing the resin in a layer-by-layer manner.

Compared to other 3D printing techniques, PμSL offers several advantages. Inkjet printing, although capable of depositing nano/microscale droplets [160, 161], cannot achieve true 3D structures at such fine scales. Stereolithography (SLA), another laser-induced localized polymerization method, also exhibits lower precision limits compared to PμSL [162]. While two-photon polymerization (2PP) can achieve even higher precision down to 100nm, it relies on a point-by-point writing process that sacrifices speed [163, 164, 165]. PμSL's high precision and relatively wide printing area make it well-suited for fabricating intricate micro-structures for applications like optical diffractive neural networks. These bio-inspired systems require precisely engineered 3D photonic architectures to manipulate light for optical computation. PμSL's ability to create such structures directly from computer-aided design (CAD) models offers a significant advantage over alternative techniques.

Our work uses the two-color PμSL system that utilizes two DMDs with different wavelengths to independently control photoinitiation and photoinhibition of polymerization in a photocurable resin. In this system, a photoinitiator absorbs blue light from one DMD to generate radicals and initiate polymerization, while a photoinhibitor absorbs UV light from the second DMD to generate inhibiting species that terminate polymerization. By carefully selecting the photoinitiator and photoinhibitor with complementary absorption spectra and optimizing their concentrations, we can control the two processes independently. The thickness of the cured polymer is determined by the intensity ratio of blue to UV light, enabling high-precision printing.

4.2 Model development for 3D printing

4.2.1 ODEs chemical model

The development of ODEs for modeling free-radical photopolymerization in 3D printing processes requires a thorough understanding of the underlying chemical reactions. As demonstrated in prior works [166, 156], these light-initiated reactions involve complex interactions between various species, such as initiators, co-initiators, monomers, and inhibitors. To comprehensively represent the photopolymerization kinetics, the kinetic model is constructed based on four primary processes: initiation, propagation, termination, and inhibition.

4.2.2 Components and Terminology

Before presenting the system of ODEs, it is essential to define the components and terminology involved in the photopolymerization process:

P_0**:** Initiator in its ground state.

P_1**:** Excited Initiator, indicating the initiator in its excited state.

P_3**:** Active growing radical, playing a role in the propagation of the reaction.

C_0**:** Co-initiator, essential for initiating the polymerization process.

M**:** Monomer, the basic building block for the polymer chain.

Q**:** Oxygen, acting as an inhibitor by quenching growing radicals.

I_0**:** Inhibitor in its ground state.

I_1**:** Excited Inhibitor, the excited state of the inhibitor.

I_2**:** Inhibitor radical, a radical form of the inhibitor.

τ_1**,** τ_2**:** Deactivation constants for the initiator and inhibitor, respectively.

σ_E**,** σ_S**:** Absorption cross-sections of the initiator and inhibitor, respectively.

k_d**,** k_p**,** k_t**,** h_t**:** Kinetic constants associated with various reactions in the photopolymerization process.

4.2.3 System of ODEs

Based on the defined components and terminology, the system of ODEs modeling the photopolymerization process is as follows:

$$\frac{dP_0}{dt} = -\sigma_E I_E(P_0 - P_1) + \tau_1 P_1, \tag{4.1}$$
$$\frac{dP_1}{dt} = \sigma_E I_E(P_0 - P_1) - \tau_1 P_1 - k_d P_1 C_0, \tag{4.2}$$
$$\frac{dC_0}{dt} = -k_d P_1 C_0, \tag{4.3}$$
$$\frac{dP_3}{dt} = k_d P_1 C_0 + r_i I_2 M - r_{kt} I_2 P_3 - k_t P_3^2 - h_t P_3 Q, \tag{4.4}$$
$$\frac{dI_0}{dt} = -\sigma_S I_S(I_0 - I_1) + \tau_2 I_1, \tag{4.5}$$
$$\frac{dI_1}{dt} = \sigma_S I_S(I_0 - I_1) - \tau_2 I_1 - r_d I_1, \tag{4.6}$$
$$\frac{dI_2}{dt} = 2r_d I_1 - r_{kt} I_2 P_3 - r_i I_2 M, \tag{4.7}$$
$$\frac{dM}{dt} = -k_p P_3 M - r_i I_2 M, \tag{4.8}$$
$$\frac{dQ}{dt} = -h_t Q P_3. \tag{4.9}$$

These equations model the interactions and kinetics of the various components involved in the photopolymerization process, capturing the essential dynamics and providing a robust framework for simulating and optimizing 3D printing processes based on free-radical photopolymerization. Detailed construction and analysis of the model can be found in [156].

4.2.4 Stiffness index for system of ODEs

The stiffness index S quantifies the stiffness of a system of ODEs, and is defined as below:

$$S = \frac{\mathrm{Re}(\lambda_{\max})}{\mathrm{Re}(\lambda_{\min})}(t_1 - t_0). \tag{4.10}$$

Here, $\lambda_{\max}$ and $\lambda_{\min}$ are eigenvalues of the Jacobian matrix with largest and smallest real parts. t_0 and t_1 are initial and final times. A large S indicates large disparity between fastest and slowest dynamics, requiring implicit solvers for stability and accuracy [167]. For the system of ODEs modeling the photopolymerization process, the Jacobian matrix is as follows:

$$\begin{bmatrix}
-I_E\sigma_E & I_E\sigma_E + \tau_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
I_E\sigma_E & -C_0k_d - I_E\sigma_E - \tau_1 & -P_1k_d & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -C_0k_d & -P_1k_d & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & C_0k_d & P_1k_d & -I_2r_{kt} - 2P_3k_t - Qh_t & 0 & 0 & Mr_i - P_3r_{kt} & I_2r_i & -P_3h_t \\
0 & 0 & 0 & 0 & -I_S\sigma_S & I_S\sigma_S + \tau_2 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & I_S\sigma_S & -I_S\sigma_S - r_d - \tau_2 & 0 & 0 & 0 \\
0 & 0 & 0 & -I_2r_{kt} & 0 & 2r_d & -Mr_i - P_3r_{kt} & -I_2r_i & 0 \\
0 & 0 & 0 & -Mk_p & 0 & 0 & -Mr_i & -I_2r_i - P_3k_p & 0 \\
0 & 0 & 0 & -Qh_t & 0 & 0 & 0 & 0 & -P_3h_t
\end{bmatrix} \tag{4.11}$$

The stiffness index with parameters estimated from experimental data is around 144, which suggests that the ODE system is mildly stiff. Ideally, an implicit ODE solver would be

preferred. However, for a mildly stiff system, we choose an adaptive explicit method for computational efficiency when later incorporating a neural network into the ODEs.

4.2.5 Numerical solver for the system of ODEs

To simulate the photopolymerization process, we employ the Dormand-Prince method, a fifth-order Runge-Kutta method for its efficiency and adaptive step-size control [168]. This method dynamically adjusts the step size to maintain the desired level of accuracy while minimizing computational effort. The system of our coupled ODEs can be expressed as a vector function $\mathbf{f}(t, \mathbf{C})$, where

$$\mathbf{C} = [P_0, P_1, C_0, P_3, I_0, I_1, I_2, M, Q]^T \tag{4.12}$$

represents the dependent variables, and t is the independent variable (time). The initial conditions are given by

$$\mathbf{C}_0 = [P_0(0), P_1(0), C_0(0), P_3(0), I_0(0), I_1(0), I_2(0), M(0), Q(0)]^T. \tag{4.13}$$

The Dormand-Prince method can be summarized in Algorithm 1. It first calculates a set of slopes ($\mathbf{k}_1$ to $\mathbf{k}_6$) at each time step based on the system of ODEs $\mathbf{f}$, the current time t, the current state vector $\mathbf{C}$, and the step size h. These slopes are used to estimate the new state vector $\mathbf{C}_{\text{new}}$ and the error vector $\mathbf{C}_{\text{err}}$ using different weighted combinations. The error vector is then used to calculate the step size scaling factor δ, which is then used to adjust the step size h for the next iteration. The step size is constrained between the minimum step size $h_{\min}$ and the maximum step size $h_{\max}$ to ensure stability and efficiency. By applying the Dormand-Prince method to the system of coupled ODEs, we can numerically simulate the photopolymerization process with high accuracy and efficiency. The adaptive step-size control ensures that the solution is computed with the desired level of accuracy while minimizing the computational cost.

4.2.6 Partial differential equations (PDEs) for diffusion

In practice, chemical components not only interact locally through the coupled ODEs, but also diffuse spatially with different rate. The general form for each chemical component, considering a two-dimensional spatial domain, would include a diffusion term ∇^2, the Laplace operator, applied to the concentration of each species. Here are the PDEs incorporating diffusion for each of the species, labeled $P_0, P_1, C_0, P_3, I_0, I_1, I_2, M$, and Q:

Algorithm 1 Adaptive Runge-Kutta Method (Dormand-Prince)

Require:

Function $\mathbf{f}(t, \mathbf{C})$: The system of ODEs
Initial condition $(t_0, \mathbf{C}_0)$: The starting time t_0 and the initial state vector $\mathbf{C}_0$
Initial step size h_0: The initial size of each time step
Error tolerance ϵ: The desired level of accuracy
Minimum step size $h_{\min}$: The lower bound for the step size
Maximum step size $h_{\max}$: The upper bound for the step size
Final time t_f: The end time of the simulation

Ensure:

Approximations of $\mathbf{C}$ at each time step
$t \leftarrow t_0$ ▷ Initialize the current time
$\mathbf{C} \leftarrow \mathbf{C}_0$ ▷ Initialize the state vector
$h \leftarrow h_0$ ▷ Initialize the step size
while $t < t_f$ **do**
    Calculate slopes $\mathbf{k}_i$ for $i = 1, \ldots, 6$ based on $\mathbf{f}$, t, $\mathbf{C}$, and h
    Estimate $\mathbf{C}_{\text{new}}$: The updated state vector using a weighted combination of $\mathbf{k}_i$
    Estimate $\mathbf{C}_{\text{err}}$: The error vector using a different weighted combination of $\mathbf{k}_i$
    $\delta \leftarrow (\epsilon h / (2\|\mathbf{C}_{\text{err}}\|))^{1/5}$ ▷ Calculate the step size scaling factor
    $h \leftarrow \min(\max(\delta h, h_{\min}), h_{\max})$ ▷ Adjust the step size
    if $t + h > t_f$ **then**
        $h \leftarrow t_f - t$ ▷ Adjust the final step size to reach t_f exactly
    end if
    $t \leftarrow t + h$ ▷ Increment the current time
    $\mathbf{C} \leftarrow \mathbf{C}_{\text{new}}$ ▷ Update the state vector
end while

$$\frac{\partial P_0}{\partial t} = k_{P_0}\nabla^2 P_0 - \sigma_E I_E(P_0 - P_1) + \tau_1 P_1, \tag{4.14}$$

$$\frac{\partial P_1}{\partial t} = k_{P_1}\nabla^2 P_1 + \sigma_E I_E(P_0 - P_1) - \tau_1 P_1 - k_d P_1 C_0, \tag{4.15}$$

$$\frac{\partial C_0}{\partial t} = k_{C_0}\nabla^2 C_0 - k_d P_1 C_0, \tag{4.16}$$

$$\frac{\partial P_3}{\partial t} = k_{P_3}\nabla^2 P_3 + k_d P_1 C_0 + r_i I_2 M - r_{kt} I_2 P_3 - k_t P_3^2 - h_t P_3 Q, \tag{4.17}$$

$$\frac{\partial I_0}{\partial t} = k_{I_0}\nabla^2 I_0 - \sigma_S I_S(I_0 - I_1) + \tau_2 I_1, \tag{4.18}$$

$$\frac{\partial I_1}{\partial t} = k_{I_1}\nabla^2 I_1 + \sigma_S I_S(I_0 - I_1) - \tau_2 I_1 - r_d I_1, \tag{4.19}$$

$$\frac{\partial I_2}{\partial t} = k_{I_2}\nabla^2 I_2 + 2r_d I_1 - r_{kt} I_2 P_3 - r_i I_2 M, \tag{4.20}$$

$$\frac{\partial M}{\partial t} = k_M \nabla^2 M - k_p P_3 M - r_i I_2 M, \tag{4.21}$$

$$\frac{\partial Q}{\partial t} = k_Q \nabla^2 Q - h_t Q P_3, \tag{4.22}$$

Where ∇^2 represents the Laplacian operator, capturing the diffusion effect in space. In two dimensions, this is $\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}$ for a function $f(x, y, t)$. k is the assumed diffusion coefficient for each chemical component.

To include diffusion processes in the modeling, we employ a numerical approach based on the finite difference method. This method discretizes the spatial domain into a grid and approximates derivatives with differences, thus transforming the continuous PDE into a system of linear equations. The diffusion equation, which is central to modeling the spread of substances in space and time, for a single component in two dimensions is given by:

$$\frac{\partial C}{\partial t} = k_{\text{diff}} \left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2} \right) + R(C, x, y, t), \tag{4.23}$$

where C represents the concentration of the substance, k_{diff} is the diffusion coefficient for a particular C, and $R(C, x, y, t)$ represents the reaction terms in the system. The spatial domain is then discretized into a grid with spacing Δx and Δy in the x and y directions, respectively. The second derivatives in the diffusion term are approximated using the central difference scheme:

$$\frac{\partial^2 C}{\partial x^2} \approx \frac{C_{i+1,j} - 2C_{i,j} + C_{i-1,j}}{\Delta x^2}, \tag{4.24}$$

$$\frac{\partial^2 C}{\partial y^2} \approx \frac{C_{i,j+1} - 2C_{i,j} + C_{i,j-1}}{\Delta y^2}, \tag{4.25}$$

where $C_{i,j}$ denotes the concentration at the grid point (i, j).

For numerical computation, the system of equations derived from discretization is represented in matrix form. The diffusion part of the PDEs can be expressed as a linear system:

$$A\mathbf{C} = \mathbf{b}, \tag{4.26}$$

where A is a sparse matrix encoding the discretized Laplace operator and the effects of boundary conditions, $\mathbf{C}$ is the vector of unknown concentrations for multiple substances, and $\mathbf{b}$ includes the reaction term. The matrix A for a single component is constructed as follows:

- Diagonal elements of A represent the negative sum of the second derivative approximations in both directions: $-k_{\text{diff}}\,(2/\Delta x^2 + 2/\Delta y^2)$, assuming isotropic diffusion.
- Off-diagonal elements corresponding to adjacent points in the x and y directions represent the diffusion into or out of each grid point: $k_{\text{diff}}/\Delta x^2$ and $k_{\text{diff}}/\Delta y^2$, respectively.

Boundary conditions are incorporated by modifying the entries of matrix A and vector $\mathbf{b}$ at the edges of the spatial domain:

- For Dirichlet boundaries, specific concentration values are imposed directly.
- For Neumann or Robin boundary conditions, the modifications reflect the gradient or flux of the substance at the boundary, influencing the diffusion process.

In our implementations, we use Robin boundary conditions (also known as convective or third-type boundary conditions) [169], which model scenarios where the flux of a substance across a boundary is proportional to the difference between the substance concentration at the boundary and the ambient concentration outside:

$$-k_{\text{diff}}\frac{\partial C}{\partial n} = h_{\text{trans}}(C - C_\infty), \tag{4.27}$$

where k_{diff} is the diffusion coefficient, $\frac{\partial C}{\partial n}$ is the derivative of concentration C normal to the boundary surface, h_{trans} is the transfer coefficient, C is the concentration at the boundary, and C_∞ is the ambient concentration outside the boundary.

This approach enables the detailed modeling of diffusion processes in a PDE framework, accounting for spatial variability, reaction kinetics, and boundary influences.

4.3 Challenges and approaches in parameter estimation

Accurate estimation of the parameters in the ODE system and the diffusion coefficients in the PDE formulation are important for reliable modeling and simulation of the photopolymerization process. However, this task presents several challenges:

- The photopolymerization process is complex, involving multiple interacting components with kinetic parameters and concentration-dependent behavior. Estimating these parameters requires extensive experimental data, which may be difficult or time-consuming to obtain. Moreover, the parameters may vary depending on the specific materials, environmental conditions, and experimental setup.

- The coupling between the ODE system and the PDE formulation complicates the parameter estimation process. The concentrations of the species in the ODE system influence the diffusion processes modeled by the PDEs, while the diffusion affects the local concentrations and reaction rates in the ODEs.
- Experimental limitations and measurement uncertainties pose challenges in parameter estimation. Typically, only the monomer concentration (M) is measured as a function of time using FTIR measurements, making it challenging to estimate the kinetic parameters in the ODE system.

To address these challenges, we propose the following approaches:

- Utilize prior knowledge of chemical parameters from the literature to incorporate into our model. This approach allows us to obtain a good model even with limited experimental data, as the optimization process becomes primarily focused on fine-tuning the model to match the observed monomer concentration profile.
- Simplify the modeling of diffusion processes by assuming that the diffusion of the monomer (M) is the most important and non-zero, while the diffusion terms for other species are ignored. This assumption is justified by the fact that the monomer concentration is the only observable in the system, while all other species are hidden variables that cannot be directly measured.

By using prior knowledge and making simplifying assumptions about diffusion, we can develop a model that effectively captures the key aspects of the photopolymerization process while working within the constraints of limited experimental data. However, it is important to acknowledge that these simplifications and assumptions may introduce some uncertainties and limitations in the model's predictive power. As more experimental data becomes available, the model can be refined and updated to incorporate more detailed information about the kinetics and diffusion behavior of the various chemical species involved in the process.

4.4 Connection to Gaussian filter with simplified diffusion modeling

We simplify the system by assuming that the diffusion of the monomer (M) is the most important and non-zero, while the diffusion terms for other species are ignored. This assumption is justified by the fact that the monomer concentration is the only observable in the system, while all other species are hidden variables that cannot be directly measured.

Under this assumption, the equation for isotropic diffusion of the monomer in two dimensions is given by:

$$\frac{\partial M}{\partial t} = k_{\text{diff}} \left(\frac{\partial^2 M}{\partial x^2} + \frac{\partial^2 M}{\partial y^2} \right), \tag{4.28}$$

where $M(x, y, t)$ represents the monomer concentration at spatial location (x, y) and time t, and k_{diff} is the diffusion coefficient. The fundamental solution to this equation for an initial delta function (point source) at the origin and time $t = 0$ is the Gaussian function:

$$G(x, y, t) = \frac{1}{4\pi k_{\text{diff}} t} \exp\left(-\frac{x^2 + y^2}{4k_{\text{diff}} t}\right). \tag{4.29}$$

For a general initial monomer concentration $M(x, y, 0)$, the solution at a later time t can be obtained by convolving the initial condition with the fundamental solution $G(x, y, t)$:

$$M(x, y, t) = (G(\cdot, \cdot, t) * M(\cdot, \cdot, 0))(x, y) = \int\int G(x - \xi, y - \eta, t) M(\xi, \eta, 0) d\xi d\eta. \tag{4.30}$$

This convolution operation is equivalent to applying a Gaussian filter to the initial monomer concentration, where the standard deviation σ of the Gaussian kernel is related to the diffusion coefficient and time by $\sigma = \sqrt{2k_{\text{diff}} t}$. Therefore, applying a Gaussian filter with standard deviation σ simulates the effect of isotropic diffusion of the monomer over time t, while neglecting the diffusion terms for other species in the system.

4.5 Conventional optimization for diffusion parameter

Given the well-established relationship between the Gaussian filter and the diffusion equation, we can formulate an optimization problem to estimate the diffusion coefficient k_{diff}. Applying a Gaussian filter with standard deviation σ to the monomer concentration is equivalent to evolving the monomer concentration under the diffusion equation for a time proportional to σ^2.

The optimization problem for estimating the diffusion coefficient can be expressed as:

$$\min_{k_{\text{diff}}} \frac{1}{N} \sum_{i=1}^{N} \sum_{x,y} \left(M_t^{(i)}(x, y) - G_\sigma(M_0^{(i)})(x, y)\right)^2, \tag{4.31}$$

where:

- $M_t^{(i)}(x, y)$ represents the monomer concentration at coordinates (x, y) after the photopolymerization process and subsequent washing, as observed in the microscope image for the i-th training sample.
- $G_\sigma(M_0^{(i)})(x, y)$ denotes the result of applying the Gaussian filter (with standard deviation σ) to the initial monomer concentration $M_0^{(i)}$, simulating the effect of diffusion over time for the i-th training sample.
- σ is related to the diffusion coefficient k_{diff} and the elapsed time t by the relation $\sigma = \sqrt{2k_{\text{diff}} t}$.
- N is the total number of training samples.

The objective of the optimization is to minimize the sum of squared differences across all pixels between the observed monomer concentration M_t and the initial monomer concentration M_0. By solving this optimization problem, we can estimate the diffusion coefficient k_{diff} that best explains the transformation from M_0 to M_t. This approach will be serving as one of the baselines in comparison to the neural modeling.

4.6 Data-efficient neural modeling

Conventional optimization techniques for estimating the diffusion parameter, such as least squares fitting, are limited by their reliance on a single free parameter and the assumption of isotropic diffusion. In reality, diffusion processes are more complex, exhibiting anisotropic behavior influenced by factors like concentration gradients, temperature fluctuations, and material heterogeneity. To improve the precision of our polymerization reaction model for designing and printing, we have integrated a data-efficient neural model. This technique is designed to account for complexities not fully captured by ordinary differential equations, such as oxygen inhibition and chemical diffusion processes, which are critical in modeling the dynamics of polymerization reactions and improving manufacturing precision.

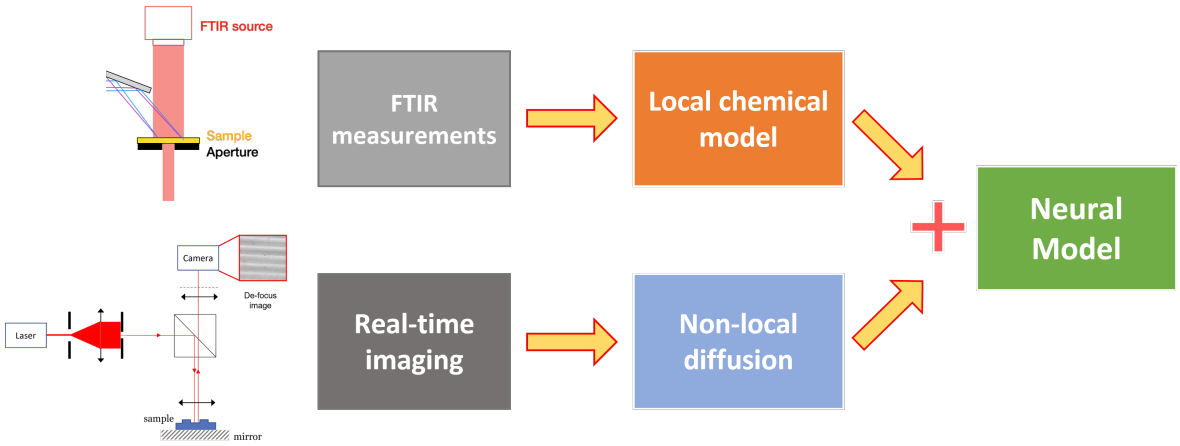


Figure 4.1: Conceptual diagram for our data-efficient neural model.

To achieve this, we incorporate a neural network term into the governing equations of the polymerization process. This term is designed to learn and represent the non-local interactions and complex dynamics that are not described by ordinary differential equations. The augmented equation for the rate of change of the monomer concentration M (which is the only observable in the experiments), is given by:

$$\frac{\partial M}{\partial t} \approx \mathbf{NN}(I_E(x,y), I_S(x,y), M(x,y), t) - k_p P_3 M - r_i I_2 M. \tag{4.32}$$

Here, the term $\mathbf{NN}\left(I_E(x,y), I_S(x,y), M(x,y), t\right)$ represents the output of a neural network model. This model is trained to encapsulate the effects of non-local interactions on the polymerization process. The inputs to the neural network are the spatial distributions of the blue wavelength intensity $I_E(x,y)$ for initiation, the UV wavelength intensity $I_S(x,y)$ for inhibition, and the monomer concentration $M(x,y)$, along with the time t. This approach allows for adapting to variations in reaction conditions and material properties that are not easily captured by conventional models with ODEs from the imaging data. A conceptual diagram for our data-efficient neural model is shown in Fig. 4.1.

4.7 Couplings between monomer and other species

Given the complete set of equations with a neural network modeling the diffusion of M:

$$\frac{\partial P_0}{\partial t} \approx -\sigma_E I_E (P_0 - P_1) + \tau_1 P_1, \tag{4.33}$$

$$\frac{\partial P_1}{\partial t} \approx \sigma_E I_E (P_0 - P_1) - \tau_1 P_1 - k_d P_1 C_0, \tag{4.34}$$

$$\frac{\partial C_0}{\partial t} \approx -k_d P_1 C_0, \tag{4.35}$$

$$\frac{\partial P_3}{\partial t} \approx k_d P_1 C_0 + r_i I_2 M - r_{kt} I_2 P_3 - k_t P_3^2 - h_t P_3 Q, \tag{4.36}$$

$$\frac{\partial I_0}{\partial t} \approx -\sigma_S I_S (I_0 - I_1) + \tau_2 I_1, \tag{4.37}$$

$$\frac{\partial I_1}{\partial t} \approx \sigma_S I_S (I_0 - I_1) - \tau_2 I_1 - r_d I_1, \tag{4.38}$$

$$\frac{\partial I_2}{\partial t} \approx 2 r_d I_1 - r_{kt} I_2 P_3 - r_i I_2 M, \tag{4.39}$$

$$\frac{\partial M}{\partial t} \approx \mathbf{NN}\left(I_E(x,y), I_S(x,y), M(x,y), t\right) - k_p P_3 M - r_i I_2 M, \tag{4.40}$$

$$\frac{\partial Q}{\partial t} \approx -h_t Q P_3. \tag{4.41}$$

The concentration of the monomer M exhibits both direct and indirect couplings with various other species in the system. The direct couplings are evident in the equations governing the dynamics of P_3 and I_2:

$$\frac{\partial P_3}{\partial t} \approx k_d P_1 C_0 + r_i I_2 M - r_{kt} I_2 P_3 - k_t P_3^2 - h_t P_3 Q, \tag{4.42}$$

$$\frac{\partial I_2}{\partial t} \approx 2 r_d I_1 - r_{kt} I_2 P_3 - r_i I_2 M, \tag{4.43}$$

$$\frac{\partial M}{\partial t} \approx \mathbf{NN}(I_E(x,y), I_S(x,y), M(x,y), t) - k_p P_3 M - r_i I_2 M. \tag{4.44}$$

The terms $r_i I_2 M$ and $-r_i I_2 M$ in the equations for P_3 and I_2, respectively, showcase the direct influence of M on these species. Similarly, the term $-k_p P_3 M$ in the equation for M highlights the direct impact of P_3 on M.

In addition to these direct couplings, M also exhibits indirect coupling with Q through P_3, as evident in their respective equations:

$$\frac{\partial Q}{\partial t} \approx -h_t Q P_3. \tag{4.45}$$

While P_0, P_1, C_0, I_0, I_1 are not coupled with M.

4.8 Overall framework

The overall framework for using the non-local neural model in the optical inverse design, depicted in Fig. 4.2, involves the following steps:

- Input DMD patterns for UV and blue light as initial mask designs.
- The neural model predicts the 3D printed structure based on the light modulation patterns.
- A gradient descent method compares the predicted structure with the intended design.
- The DMD patterns are adjusted to minimize the discrepancy between predicted and intended structures.
- The process iterates with the neural model providing updated predictions until the predicted structure aligns with the intended design.

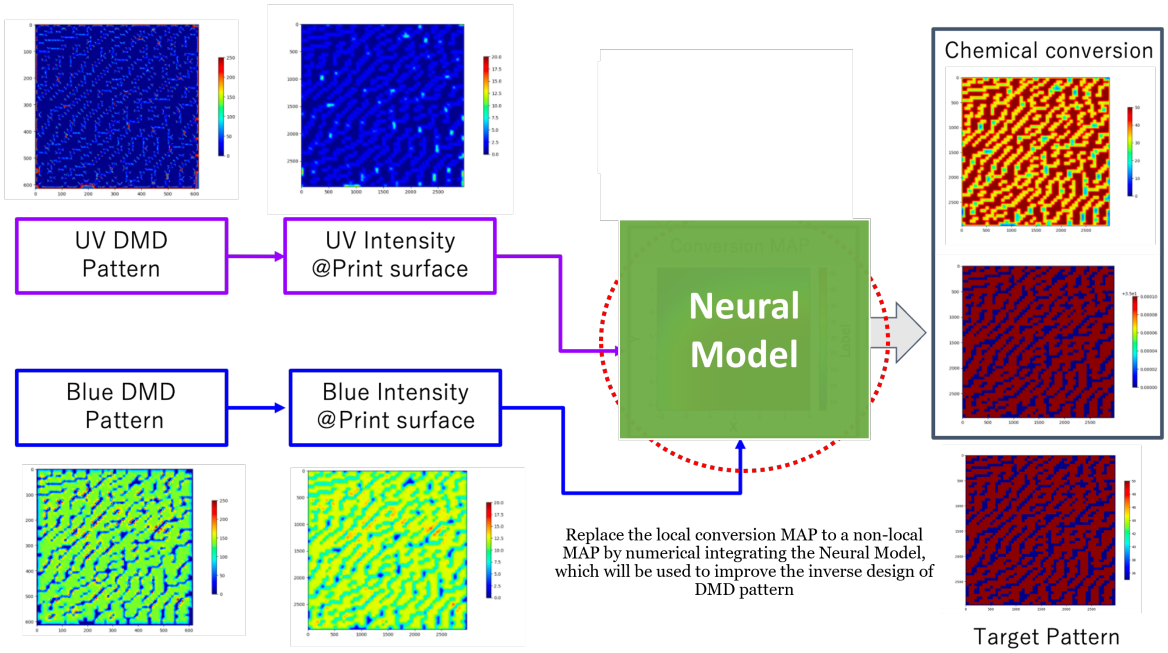
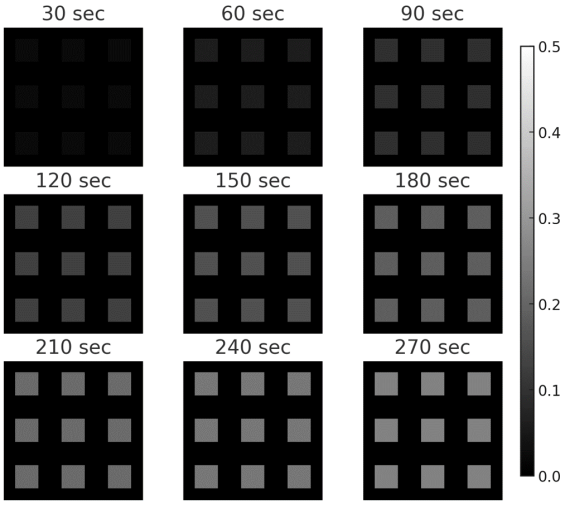


Figure 4.2: Using non-local neural model in the optical inverse design.

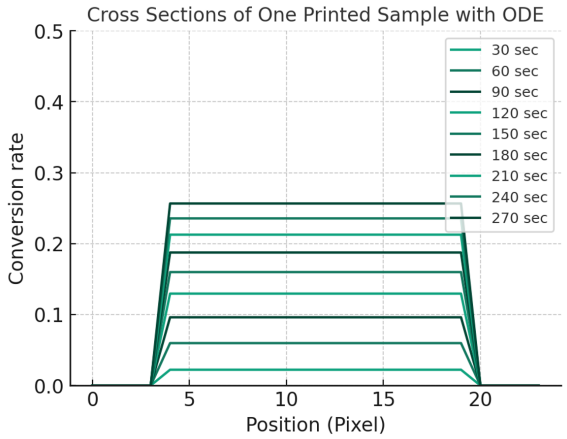
By integrating the data-efficient neural model into the inverse design process, we can predict the complex interaction between light modulation and material response, enabling the fabrication of structures that closely match the intended design.

4.9 Numerical simulations

We first compare the simulation results of ODEs and PDEs in modeling the photopolymerization process for 3D printing. Fig. 4.3 shows the ODE simulations, which model local

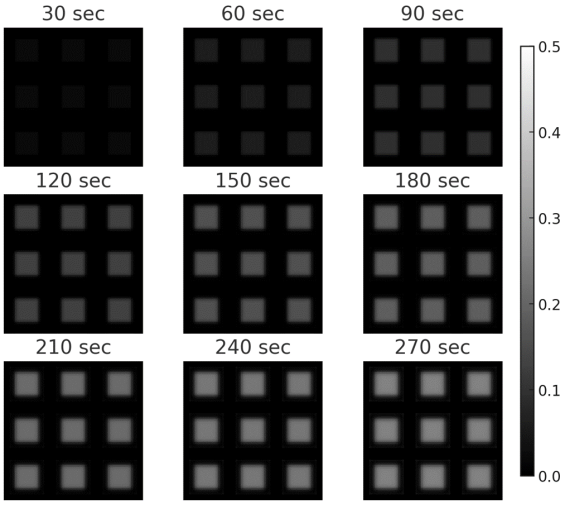


(a) Simulated conversion using ODEs

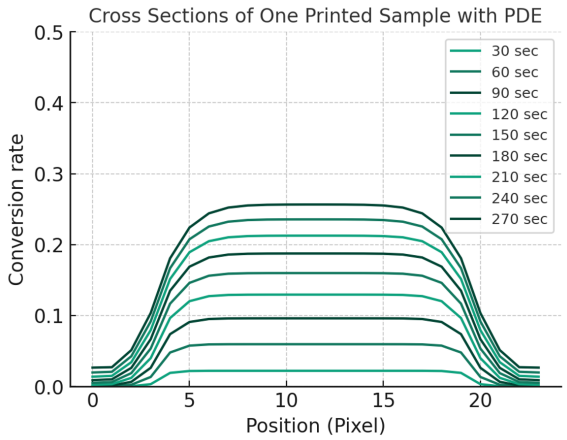


(b) Cross-section of one printed sample using ODEs

Figure 4.3: 3 x 3 printed sample simulated with ODEs, illustrating local conversion effects.



(a) Simulated conversion using PDEs



(b) Cross-section of one printed sample using PDEs

Figure 4.4: 3 x 3 printed sample simulated with PDEs, highlighting non-local conversion effects.

interactions within the polymerization process. The results exhibit a distinct and sharp conversion pattern, reflecting the localized nature of the reactions modeled by the ODEs. In contrast, Fig. 4.4 presents the PDE simulations with an assumed diffusion parameter for all the chemical species. The visual results display a more blurred conversion pattern and a smoother, more realistic cross-section texture compared to the ODE-based simulation.

The ODE and PDE simulations highlight the impact of mathematical modeling choices on the performance of 3D printed functional samples. While ODEs provide a localized reaction representation, PDEs offer a more comprehensive simulation by accounting for non-local interactions, better capturing the nuances of photopolymerization. However, it is important to note that PDEs are computationally demanding and data-intensive due to the inclusion of non-local interactions, which increases model complexity. In practice, parameter estimation for diffusion parameters in PDEs poses challenges due to the need for 2D spatial and temporal data, computationally intensive forward processes, complex optimization. To address these challenges, our data-efficient neural model serves as a surrogate model trained with imaging data. By approximating the forward process with a computationally cheaper neural network, we can efficiently perform parameter estimation and reduce the overall computational cost while maintaining accuracy.

4.10 Capturing real-time dynamics by coherent diffractive imaging

Coherent diffractive imaging (CDI) is a powerful, non-destructive imaging technique [170]. It is uniquely suited for real-time imaging during 3D printing because it does not require lenses, which could interfere with the printing process. In contrast, lens-based imaging methods, such as optical microscopy [171, 172], require focusing optics and have a limited depth of field, making them unsuitable for capturing sharp images of objects with large height variations. CDI can obtain high-quality reconstructions with a wide field of view, allowing for the observation of both fine details and larger-scale structures in real-time, regardless of the object's height.

Fig. 4.5 illustrates how we use CDI in reflection mode to measure the printing dynamics in real-time. Note that the printing optics is orthogonal to the CDI optics and is not shown in the figure.

The scattering model for diffractive imaging is based on non-paraxial version of beam propagation method (BPM) [173, 174, 93], where the exiting wave ψ after interacting with an object under thin-film approximation is

$$\psi(x,y) = \mathscr{F}^{-1}\left[\mathscr{F}[p(x,y)f(x,y)]e^{-i(k_0-\sqrt{k_0^2-k_x^2-k_y^2})\delta z}\right]. \tag{4.46}$$

Here, $p(x,y)$ is the illumination profile propagating along the z axis, $f(x,y)$ is the phase-only scattering object, δz is the layer thickness of the object, k_0 is the wavenumber in the object medium, $\mathscr{F}$ is the two-dimensional Fourier transform. The intensity of the exiting wave is measured in a de-focused plane, and is

$$g(x,y) = \left|\mathscr{F}^{-1}[\mathscr{F}[\psi(x,y)]e^{-i(k_1-\sqrt{k_1^2-k_x^2-k_y^2})\Delta z}]\right|^2 + \mathscr{N}, \tag{4.47}$$

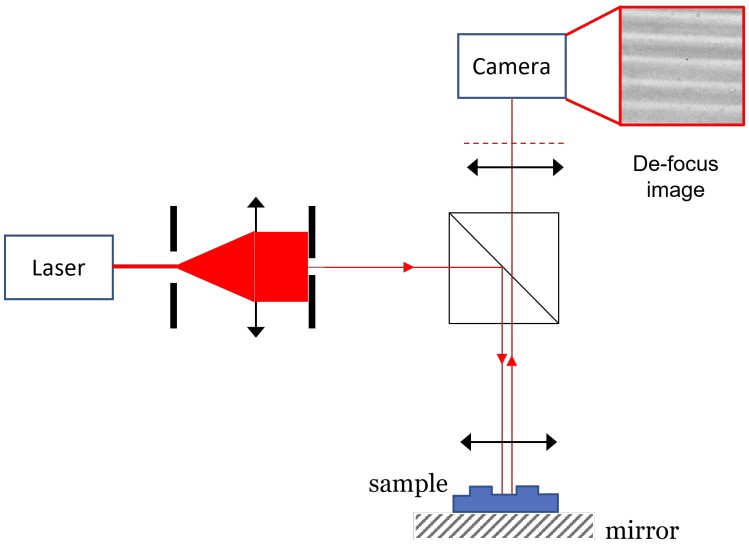


Figure 4.5: Real-time CDI measurements. The probe beam is in 632nm wavelength illuminating the sample. After reflected from the mirror and pass once again across the sample, the exisitng waving is measured with a CCD camera de-focused to the top sufrace of the printing sample.

where k_1 is the wavenumber in the background medium, Δz is the de-focus displacement, $\mathscr{N}$ is the additive background noise.

To estimate the optimal Δz for the de-focus displacement, we can consider the phase transfer function in the Fresnel region, which is proportional to

$$\sin\left[\pi\lambda(\xi^2+\eta^2)z\right], \tag{4.48}$$

Another constraint is the diffraction limit $\frac{\mathrm{NA}}{\lambda}$, which sets the highest frequency that can be recovered. For our system, $\lambda = 632$nm, we have the phase transfer function as shown in Fig. 4.6 for four defocus distances.

By using CDI's ability to capture real-time data during the printing process, we can obtain valuable datasets for pre-training our data-efficient neural model. These datasets provide crucial insights into the complex dynamics underlying the fabrication process, enabling the neural model to learn and predict the behavior of the system.

4.11 Training the neural model

To train the data-efficient neural model for the inverse design, we employ a two-stage approach: pre-training with real-time CDI data and fine-tuning with microscope images of the final printed patterns.

In the first stage, we collect real-time diffraction patterns using CDI during the photopolymerization process (Fig. 4.7). These diffraction patterns capture the dynamic evolution of the printed structure. We then apply a gradient descent algorithm to reconstruct the real-time CDI data, providing a time-series of 2D spatial images that represent the intermediate states of the printed pattern (Fig. 4.8).

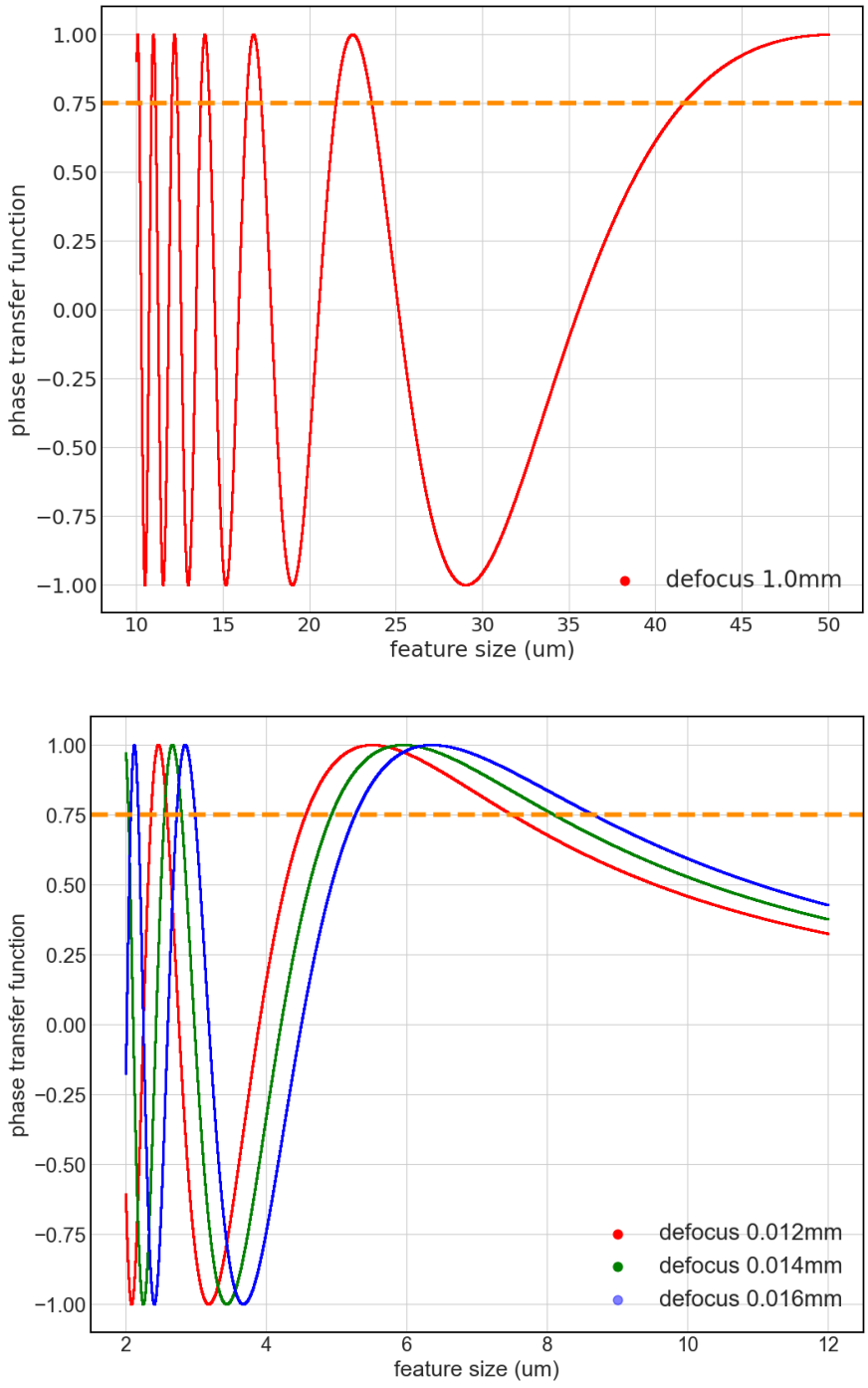


Figure 4.6: Numerical plot for the phase transfer function at four defocus distances. The orange bar represents phase transfer function with value at 0.75.

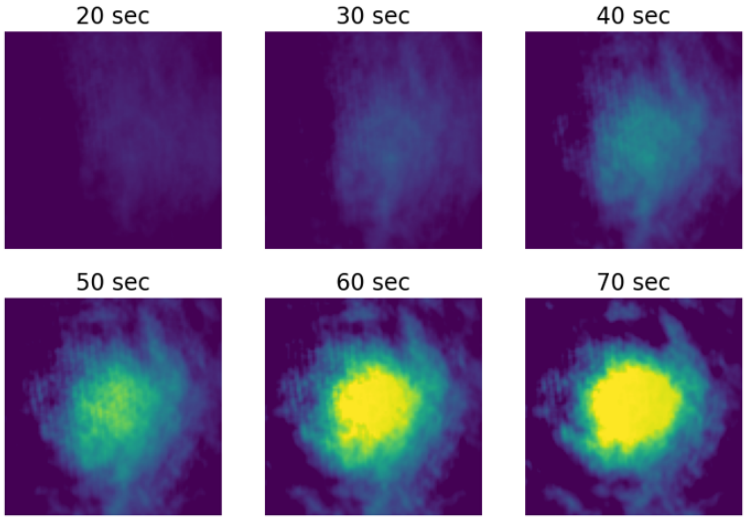


Figure 4.7: Real-time CDI reconstructions. Experimental settings: Blue intensity - 2500mV, UV intensity - 0mV, Exposure time - 70 sec, Pattern size - approximately 5um x 5um x 1um.

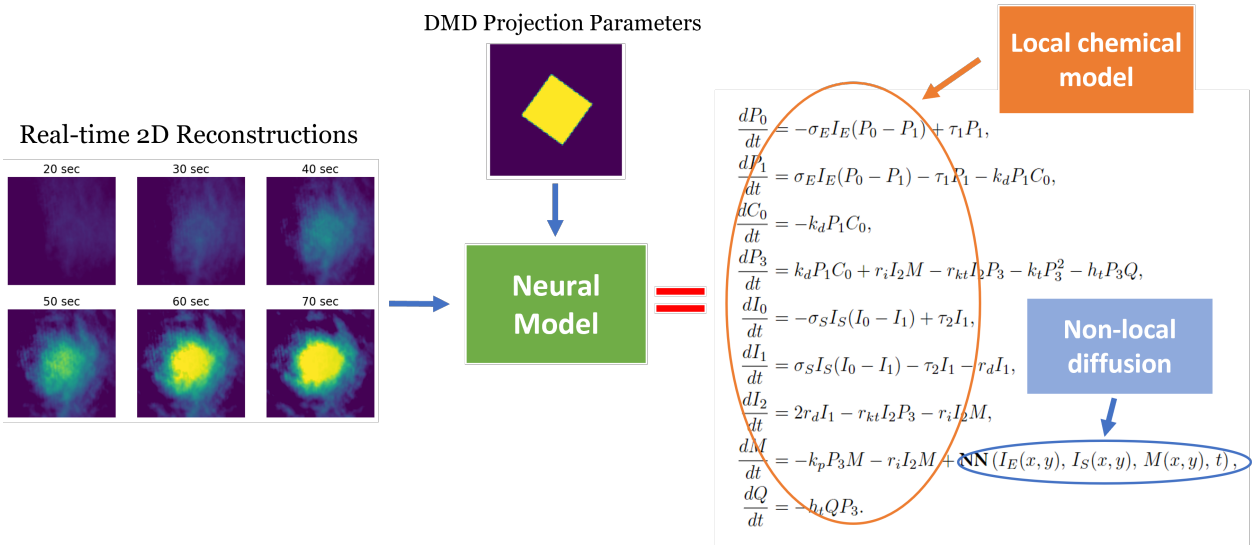


Figure 4.8: Pre-training CDI reconstructions.

The pre-training of the neural model is performed using the reconstructed CDI data. The neural network is incorporated into the ODE equations along with the other relevant ODEs, and the entire ODE system is integrated using the odeint function with the Dormand-Prince method. The objective is to train the model to predict the CDI reconstruction at each time step, given the DMD patterns for blue and UV light, and time. The optimization objective for pre-training the neural model can be expressed as follows:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\| \hat{x}_t^{(i)} - \text{odeint}\left(\frac{d\mathbf{C}}{dt}, \mathbf{C}_0^{(i)}, t, \text{args} = (I_E^{(i)}(x,y), I_S^{(i)}(x,y), \mathbf{NN}_\theta) \right)_t \right\|_1 \tag{4.49}$$

where:

- θ represents the parameters of the neural model,
- N is the total number of training samples,
- T is the total number of time steps,
- $\hat{x}_t^{(i)}$ is the target CDI reconstruction at time step t for the i-th training sample,
- $\frac{d\mathbf{C}}{dt}$ represents the entire ODE system, which includes the neural network $\mathbf{NN}_\theta$ and other relevant ODEs,
- $\mathbf{C}_0^{(i)}$ is the initial condition for all the variables in the ODE system for the i-th training sample,
- odeint$(\cdot)$ is the function that integrates the ODE system using the Dormand-Prince method, taking as input the ODE system $\frac{d\mathbf{C}}{dt}$, initial condition $\mathbf{C}_0^{(i)}$, time points t, and the DMD patterns for blue and UV light, $I_E^{(i)}(x,y)$ and $I_S^{(i)}(x,y)$, respectively, as well as the neural network $\mathbf{NN}_\theta$ for modeling the diffusion of monomer M,
- $\|\cdot\|_1$ denotes the L1 norm, which measures the absolute difference between the predicted and target CDI reconstructions.

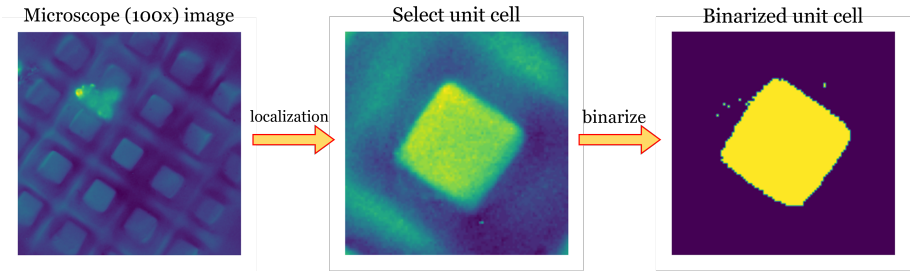


Figure 4.9: Data pre-process for microscope images. Experimental settings: Blue intensity - 2500mV, UV intensity - 0mV, Exposure time - 90 sec, Pattern size - approximately 5.63um x 5.63um x 1um.

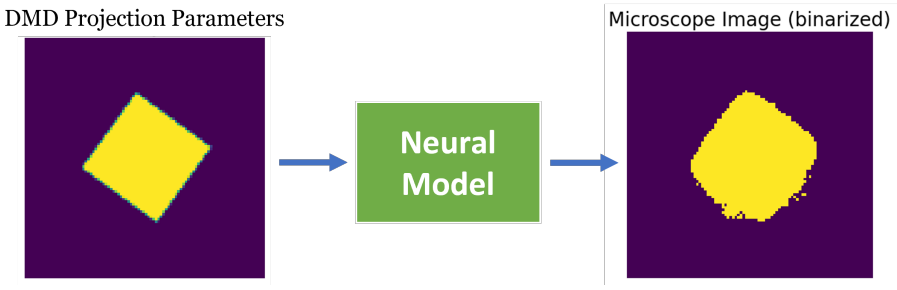


Figure 4.10: Fintuning with microscope images.

In the second stage, we collect microscope images of the final printed patterns after the completion of the photopolymerization process and subsequent washing (Fig. 4.9). These images represent the actual outcome of the printing process and provide ground truth data for fine-tuning the neural model.

The fine-tuning step involves continual training the pre-trained model using the microscope images. The input to the model consists of the DMD projection parameters and time, while the output is the predicted printed features (Fig. 4.10). The optimization objective for fine-tuning the neural model can be expressed as follows:

$$\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \left\| m^{(i)} - \text{odeint} \left(\frac{d\mathbf{C}}{dt}, \mathbf{C}_0^{(i)}, t, \text{args} = (I_E^{(i)}(x,y), I_S^{(i)}(x,y), \mathbf{NN}_\theta) \right)_{t=T} \right\|_1 \tag{4.50}$$

where:

- $m^{(i)}$ is the microscope image of the final printed pattern for the i-th training sample,
- $t = T$ denotes the final time step of the integration, corresponding to the completion of the photopolymerization process.

By combining the pre-training and the fine-tuning stages, the neural model learns to capture both the dynamic evolution of the photopolymerization process and the final outcome of the printed patterns. This two-stage training approach enables the model to effectively learn the complex relationships between DMD patterns, time, and the resulting printed structures, ultimately facilitating the inverse design of DMD masks for targeted 3D printing applications.

4.12 Experimental results

To validate the performance of the proposed data-efficient neural model, we conduct a series of experiments using our optical printing and imaging system. Fig. 4.11 showcases a representative set of data collected from the system. From left to right, we have: (a) the DMD pattern used for printing, (b) the diffractive pattern obtained from raw CDI measurements during the printing process, (c) the CDI reconstruction obtained by applying an iterative

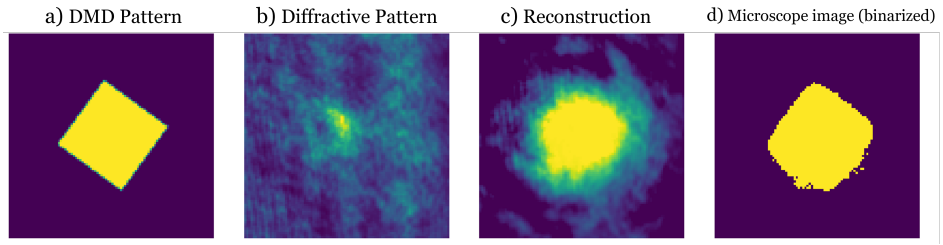


Figure 4.11: Example data collected from the optical printing and imaging system.

algorithm to the diffractive pattern, and (d) the binarized microscope image of the final printed structure.

Fig. 4.12 provides an illustrative overview of where and how we collect data from the optical printing and imaging system. The DMD projector is used to display the input pattern, which is then projected onto the photopolymer resin. During the printing process, the CDI camera captures the diffractive patterns generated by the interaction of light with the evolving structure. After printing, the sample is washed and imaged using a microscope to obtain the final printed pattern.

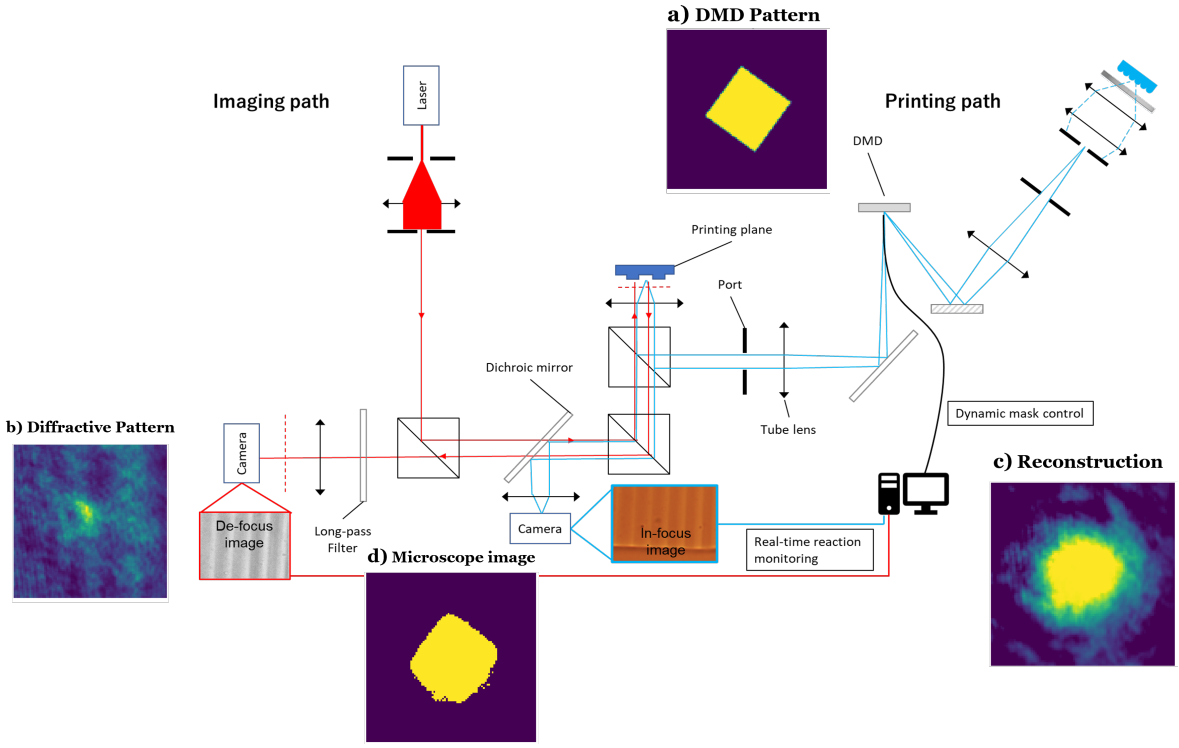


Figure 4.12: Illustrative figure on where and how we collect the data from the optical printing and imaging system.

Fig. 4.13 compares the predictions of the neural model, a conventional local chemical model (DMD), a diffusion model (DMD diffused), and the actual printed patterns (microscope). The neural model predictions (top two rows) are shown for different numbers of

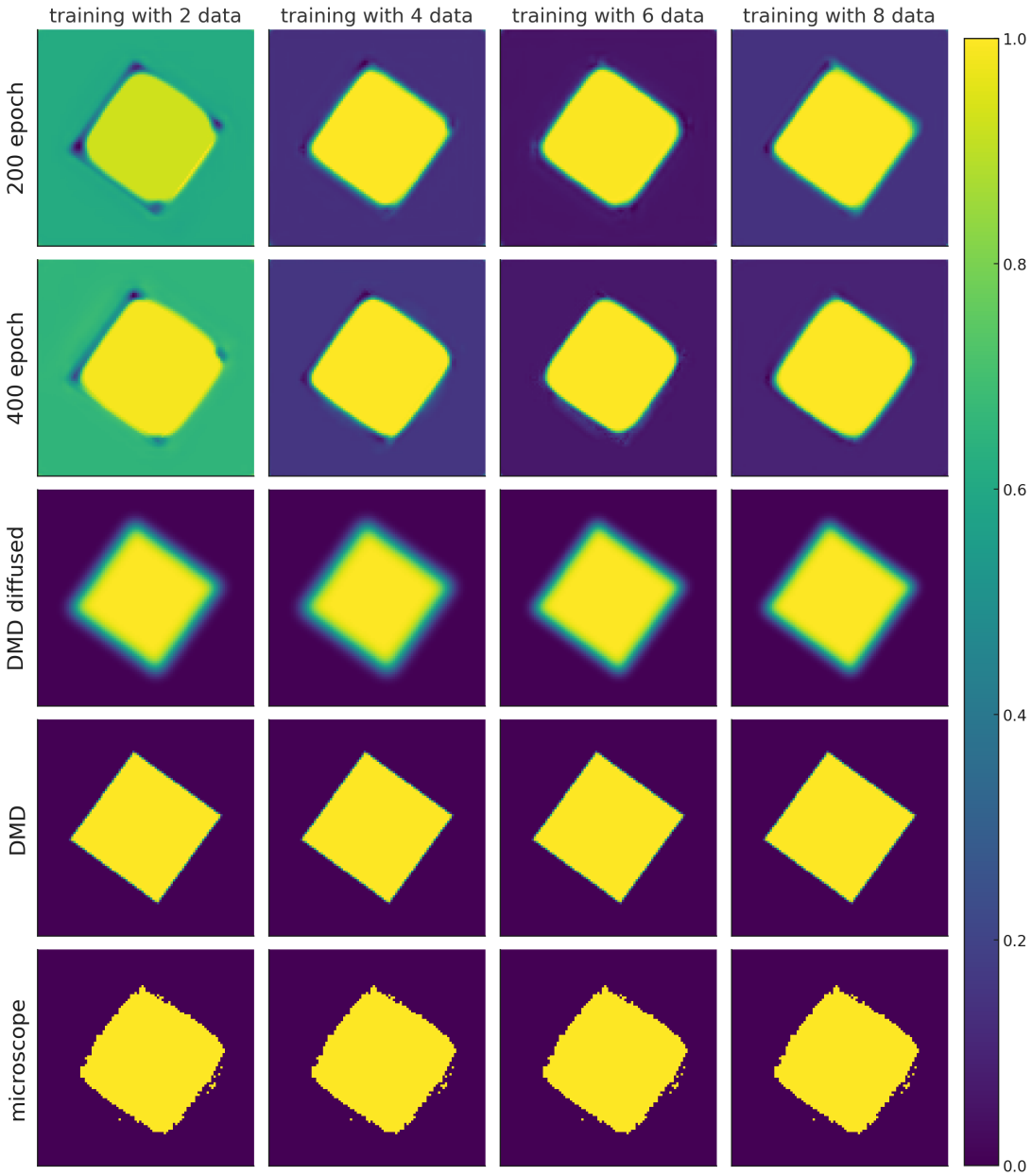


Figure 4.13: Visual comparison between neural model predictions and local chemical model predictions. The top-2 rows represent the neural model trained with different epochs, DMD diffused represents the predictions from conventional optimization for the diffusion parameter. DMD represents the local chemical model prediction for the pattern, and the microscope is the actual pattern after printing. Each column for the top-2 rows represent neural model trained with different data-point. The pattern size is about 5.63um x 5.63um x 1um physically, where 1um is the layer thickness.

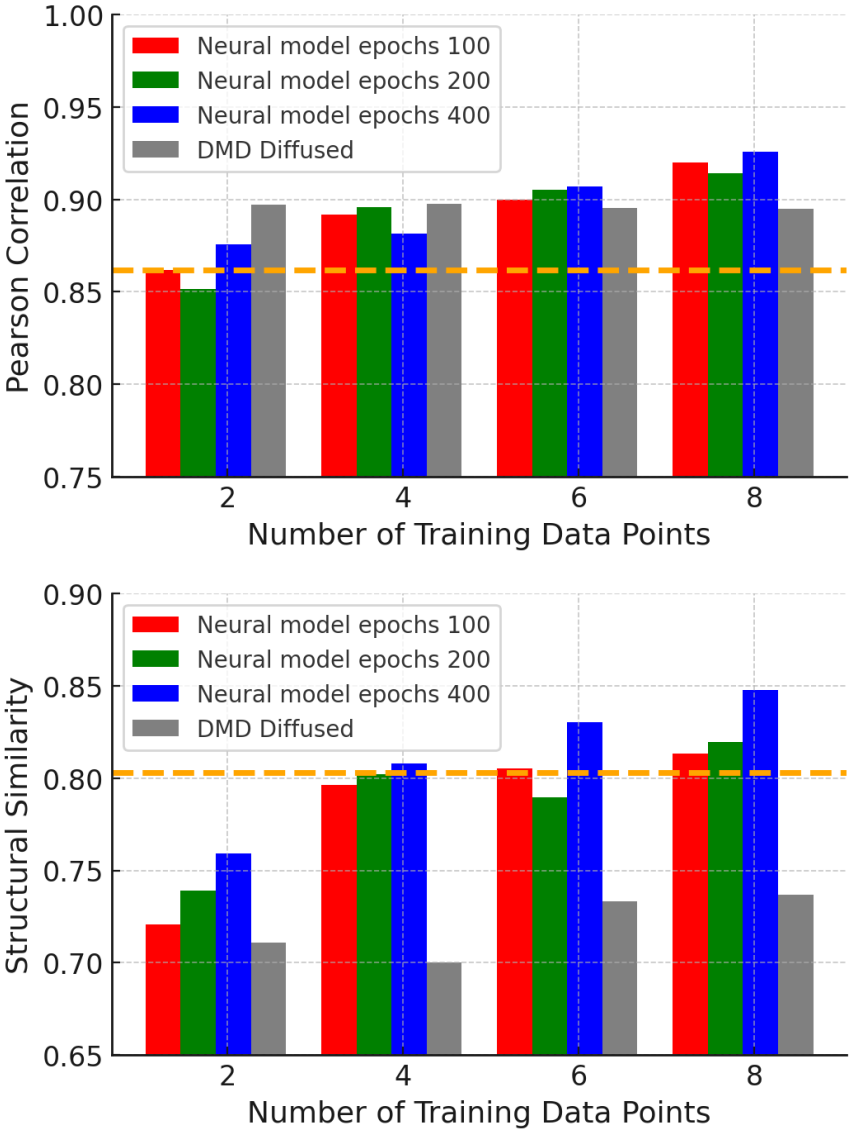


Figure 4.14: Quantitative comparison between neural model trained with different data points with pearson correlation and structural similarity metrics (higher the better). The orange dotted line represents the accuracy for local chemical model. Though 4 data-points would be enough for the neural model outperforms local chemical model, more training data further improves the results.

training epochs and data points, while the DMD diffused predictions (third row) are shown for different numbers of data points. The DMD diffused model relies on a single free parameter, resulting in predictions that appear blurred and lack the intricate details present in the actual printed patterns. In contrast, the neural model's predictions improve with increasing training data and epochs, capturing more of the complex features found in the microscope images. The local chemical model (DMD, fourth row) provides a basic prediction of the pattern, but it fails to account for the diffusion effects and intricate details observed in the actual printed patterns (microscope, bottom row). From the visual comparison, we observe that the neural model predictions closely resemble the actual printed patterns, capturing fine details and intricate structures that the local chemical model often fails to reproduce, particularly around the edges. This demonstrates the neural model's ability to learn and predict the complex, nonlinear dynamics of the photopolymerization process from a larger dataset, outperforming the conventional local chemical model by modeling the intricate relationships between various factors influencing the polymerization process.

To quantitatively assess the performance of the neural model, we use two metrics: Pearson correlation and structural similarity (SSIM). Fig. 4.14 presents a comparison of the neural model and DMD diffused model trained with different numbers of data points, along with the accuracy of the local chemical model (indicated by the orange dotted line). The results show that the neural model consistently outperforms the local chemical model, even when trained with as few as four data points. Moreover, increasing the number of training data points further improves the performance of the neural model, as evidenced by the higher Pearson correlation and SSIM scores. The DMD diffused model, indicated by the grey color bars, lags behind the neural model in SSIM but exhibits comparable PCC performance. Unlike the neural model, the DMD diffused model's performance does not improve with increasing training data points, which is not surprising given that it has only one diffusion parameter in the optimization.

The experimental results demonstrate the effectiveness of the proposed data-efficient neural model in predicting the outcomes of the photopolymerization process. By learning from a limited number of CDI measurements and microscope images, the neural model can accurately predict the final printed patterns, surpassing the performance of the conventional diffusion model and local chemical model. This highlights the potential of the neural model for efficient inverse design of DMD masks, enabling the creation of complex, high-precision 3D structures with reduced experimental overhead and computational complexity.

4.13 Limitations and discussion

Our data-efficient neural modeling approach for high-precision 3D printing using two-color projection micro-stereolithography (PμSL) has several key limitations that should be considered:

1. Performance of the neural model may be affected by the spacing between printed features. As feature spacing decreases, more pronounced non-local diffusion effects can occur due to increased interactions between nearby regions during the photopolymerization process. This can lead to potential challenges in predicting the final printed patterns, especially for designs with closely spaced and intricate features.

2. The prediction accuracy of our approach is also limited by the quality of the real-time CDI data and post-printing microscope images used for training the neural network. Higher quality and lower noise in these imaging modalities can improve the model's ability to capture fine details and subtle variations in the structures.

3. Generalization of the trained neural model to new structures or materials not represented in the training data may be restricted. The diversity of the training dataset play a crucial role in determining the model's ability to predict the outcomes for new designs or material compositions. Limited training data diversity can lead to overfitting and reduced performance on out-of-distribution samples.

Despite these limitations, our approach offers benefits over conventional methods in terms of prediction accuracy and data efficiency. Future research directions could focus on addressing the mentioned limitations by exploring techniques such as transfer learning and domain adaptation to improve the generalization and robustness of the neural model across a wider range of structures and materials.

4.14 Summary and contributions

In this chapter, we present a data-efficient neural modeling approach for high-precision 3D printing using two-color PμSL. Our key contributions can be summarized as follows:

1. We develop a neural model that augments traditional ODEs to effectively capture non-local diffusion effects and complex dynamics inherent in the photopolymerization process, enabling accurate predictions of the printing outcomes.

2. We propose a two-stage training approach that combines both real-time CDI data and post-printing microscope images. In the first stage, the model learns to predict the printing dynamics based on the CDI data, capturing the temporal evolution of the structures. In the second stage, the model is fine-tuned using the microscope images to predict the final printed patterns. This two-stage approach allows our model to effectively learn both the dynamic process and the final outcome.

3. Through experimental validations, we demonstrate strong performance of our neural model compared to conventional modeling methods. Our approach achieves higher prediction accuracy and requires a few training data, making it particularly advantageous in data-limited scenarios. This data efficiency is important for practical applications where obtaining large amounts of experimental data may be time-consuming or costly.

Our data-efficient neural model serves as a valuable tool for creating complex, high-precision 3D structures in scenarios where limited training data is available. Its efficiency and predictive power have the potential to drive breakthroughs in next-generation optical computing and other advanced applications.

Chapter 5

Conclusion

In this thesis, we have explored the potential of data-efficient machine learning in advancing computational imaging techniques. By intelligently integrating physical models with the adaptability of deep learning, we have demonstrated strong improvements in imaging accuracy, efficiency, and noise resilience under stringent data and time constraints.

For X-ray tomography, we developed a Physics-assisted Generative Adversarial Network (PGAN) for limited-angle, low-photon imaging conditions. Our PGAN integrates a maximum-likelihood estimate informed by known physics to address the challenges of ill-conditioning and Poisson shot noise inherent in such scenarios. The separation of physical constraints from the network optimization proved to yield better reconstruction quality compared to less physics-assisting strategies. Using the CircuitFaker model to generate 3D circuit objects, we demonstrated that PGAN improves reconstruction quality under limited viewing angles and photon flux compared to traditional maximum-likelihood estimates and alternative deep learning approaches.

In addition, we introduced a noise-resilient deep-reconstruction algorithm for X-ray tomography of integrated circuits that incorporates maximum *a posteriori* (MAP) reconstructions into the neural network input. By integrating a Gaussian noise prior and a sparsity-promoting prior into MAP reconstructions, we mitigate input distribution shifts from varying noise levels, improving learned prior noise resilience without additional noise statistic sampling. This data-efficient approach eliminates the need to collect and label data under different noise conditions, making it particularly advantageous in scenarios like circuit imaging where obtaining varied-noise-level training datasets is difficult. Our method achieves high fidelity reconstructions requiring fewer photons compared to prior works with FBP+UNet approach, with up to an 8× reduction in required photons in simulations and a 2.5× reduction in experiments.

Lastly, we introduced a data-efficient neural model that captures the diffusion processes underlying high-precision 3D printing. The model combines real-time phase imaging, conventional camera imaging, and pre-determined system of ordinary differential equations (ODEs). This approach demonstrates improved accuracy in predicting the printing of optical diffractive elements, which is an important step towards more advanced and efficient optical neural computing systems. The data-efficient nature of our neural model reduces experimental overhead and computational complexity in the inverse design for 3D printing, serving as a valuable tool for creating complex, high-precision 3D structures, even in data-limited sce-

narios.

These advancements highlight the potential of data-efficient machine learning in pushing the boundaries of computational imaging. By harnessing the complementary strengths of physical models and deep learning, we can unlock new possibilities for applications demanding high-fidelity imaging under challenging conditions. As we continue to refine and expand upon these techniques, we anticipate a profound impact on fields ranging from biomedical diagnostics and materials science to additive manufacturing and optical computing, enabling faster, more accurate, and more efficient imaging and analysis of complex systems.

References

[1] Eric H Lee et al. "Discovery through the computational microscope". In: *Structure* 17.10 (2009), pp. 1295–1306.

[2] Avinash C Kak and Malcolm Slaney. *Principles of computerized tomographic imaging*. SIAM, 2001.

[3] Gary H Glover. "Overview of functional magnetic resonance imaging". In: *Neurosurgery Clinics* 22.2 (2011), pp. 133–139.

[4] Peter NT Wells. "Ultrasound imaging". In: *Physics in medicine & biology* 51.13 (2006), R83.

[5] Atsushi Momose et al. "Phase-contrast X-ray computed tomography for observing biological soft tissues". In: *Nature Medicine* 2.4 (1996), pp. 473–475.

[6] Roland Eils and Chaitanya Athale. "Computational imaging in cell biology". In: *The Journal of cell biology* 161.3 (2003), pp. 477–481.

[7] Carol Muehleman et al. "Diffraction-enhanced imaging of musculoskeletal tissues using a conventional X-ray tube". In: *Academic Radiology* 16.8 (2009), pp. 918–923.

[8] Luc Salvo et al. "3D imaging in material science: Application of X-ray tomography". In: *Comptes Rendus Physique* 11.9-10 (2010), pp. 641–649.

[9] Clément YJ Hémonnot and Sarah Köster. "Imaging of biological materials and cells by X-ray scattering and diffraction". In: *ACS Nano* 11.9 (2017), pp. 8542–8559.

[10] Mahbub Alam et al. "Impact of X-ray tomography on the reliability of integrated circuits". In: *IEEE Transactions on Device and Materials Reliability* 17.1 (2017), pp. 59–68.

[11] Joseph N Mait, Gary W Euliss, and Ravindra A Athale. "Computational imaging". In: *Advances in Optics and Photonics* 10.2 (2018), pp. 409–483.

[12] Mirko Holler et al. "Three-dimensional imaging of integrated circuits with macro to nanoscale zoom". In: *Nature Electronics* 2.10 (Aug. 2019), pp. 464–470. ISSN: 2520-1131. DOI: 10.1038/s41928-019-0309-z. URL: https://doi.org/10.1038/s41928-019-0309-z.

[13] Mark E Davison. "The ill-conditioned nature of the limited angle tomography problem". In: *SIAM Journal on Applied Mathematics* 43.2 (1983), pp. 428–448.

[14] Alfred K Louis. "Incomplete data problems in X-ray computerized tomography". In: *Numerische Mathematik* 48.3 (1986), pp. 251–262.

[15] Ge Wang et al. *Machine learning for tomographic imaging*. IOP Publishing, 2019.

[16] Erich Kobler et al. "Variational deep learning for low-dose computed tomography". In: *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE. 2018, pp. 6687–6691.

[17] Mauricio Araya-Polo et al. "Deep-learning tomography". In: *The Leading Edge* 37.1 (2018), pp. 58–66.

[18] Tatiana A Bubba et al. "Learning the invisible: A hybrid deep learning-shearlet framework for limited angle computed tomography". In: *Inverse Problems* 35.6 (2019), p. 064002.

[19] Yixing Huang et al. "Data consistent artifact reduction for limited angle tomography with deep learning prior". In: *International Workshop on Machine Learning for Medical Image Reconstruction*. Springer. 2019, pp. 101–112.

[20] Jiaxi Wang et al. "Deep learning based image reconstruction algorithm for limited-angle translational computed tomography". In: *PLOS One* 15.1 (2020), e0226963.

[21] Yixing Huang et al. "Limited angle tomography for transmission X-ray microscopy using deep learning". In: *Journal of Synchrotron Radiation* 27.2 (2020), pp. 477–485.

[22] Vegard Antun et al. "On instabilities of deep learning in image reconstruction and the potential costs of AI". In: *Proceedings of the National Academy of Sciences* 117.48 (2020), pp. 30088–30095.

[23] Nina M Gottschling et al. "The troublesome kernel: why deep learning for inverse problems is typically unstable". In: *arXiv:2001.01258* (2020).

[24] Alexandre Goy et al. "Low photon count phase retrieval using deep learning". In: *Physical Review Letters* 121.24 (2018), p. 243902.

[25] Yu Sun, Zhihao Xia, and Ulugbek S Kamilov. "Efficient and accurate inversion of multiple scattering with deep learning". In: *Optics Express* 26.11 (2018), pp. 14678–14688.

[26] Iksung Kang, Fucai Zhang, and George Barbastathis. "Phase extraction neural network (PhENN) with coherent modulation imaging (CMI) for phase retrieval at low photon counts". In: *Optics Express* 28.15 (2020), pp. 21578–21600.

[27] Zhen Guo and Abraham Levitan. *Randomized probe imaging through deep k-learning*. Version 1.0.0. Nov. 2021. URL: https://github.com/zguo0525/Randomized-probe-imaging-through-deep-k-learning.

[28] Zhen Guo et al. "Physics-assisted generative adversarial network for X-ray tomography". In: *Optics Express* 30.13 (2022), pp. 23238–23259.

[29] G. Barbastathis. *Lecture Notes for 2.c27/2.c67 Computational Imaging: Physics and Algorithms*. Lecture notes, Massachusetts Institute of Technology. Distribution restricted to the class students. 2023.

[30] Stanley Osher et al. "An iterative regularization method for total variation-based image restoration". In: *Multiscale Modeling & Simulation* 4.2 (2005), pp. 460–489.

[31] Tom Goldstein and Stanley Osher. "The split Bregman method for L1-regularized problems". In: *SIAM journal on imaging sciences* 2.2 (2009), pp. 323–343.

[32] Sebastian Ruder. "An overview of gradient descent optimization algorithms". In: *arXiv preprint arXiv:1609.04747* (2016).

[33] John A Scales. "Tomographic inversion via the conjugate gradient method". In: *Geophysics* 52.2 (1987), pp. 179–185.

[34] Todd K Moon. "The expectation-maximization algorithm". In: *IEEE Signal processing magazine* 13.6 (1996), pp. 47–60.

[35] Ayan Sinha et al. "Lensless computational imaging through deep learning". In: *Optica* 4.9 (2017), pp. 1117–1125.

[36] Christopher Metzler et al. "prDeep: Robust Phase Retrieval with a Flexible Deep Network". In: *Proceedings of the 35th International Conference on Machine Learning*. Ed. by Jennifer Dy and Andreas Krause. Vol. 80. Proceedings of Machine Learning Research. PMLR, July 2018, pp. 3501–3510. URL: https://proceedings.mlr.press/v80/metzler18a.html.

[37] Mo Deng et al. "Learning to synthesize: Robust phase retrieval at low photon counts". In: *Light: Science & Applications* 9.1 (2020), pp. 1–16.

[38] Yuhe Zhang et al. "PhaseGAN: A deep-learning phase-retrieval approach for unpaired datasets". In: *arXiv preprint arXiv:2011.08660* (2020).

[39] Mo Deng et al. "On the interplay between physical and content priors in deep learning for computational imaging". In: *Opt. Express* 28.16 (2020), pp. 24152–24170.

[40] Iksung Kang, Alexandre Goy, and George Barbastathis. "Dynamical machine learning volumetric reconstruction of objects' interiors from limited angular views". In: *Light: Science & Applications* 10.1 (2021), pp. 1–21.

[41] Iksung Kang et al. "Attentional Ptycho-Tomography (APT) for three-dimensional nanoscale X-ray imaging with minimal data acquisition and computation time". In: *arXiv preprint arXiv:2212.00014* (2022).

[42] Bo Zhu et al. "Image reconstruction by domain-transform manifold learning". In: *Nature* 555.7697 (2018), pp. 487–492.

[43] Ge Wang et al. "Image reconstruction is a new frontier of machine learning". In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1289–1296.

[44] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60.6 (2017), pp. 84–90.

[45] Yan Yang et al. "ADMM-CSNet: A deep learning approach for image compressive sensing". In: *IEEE transactions on pattern analysis and machine intelligence* 42.3 (2018), pp. 521–538.

[46] Yair Rivenson et al. “Phase recovery and holographic image reconstruction using deep learning in neural networks”. In: *Light: Science & Applications* 7.2 (2018), pp. 17141–17141.

[47] Guang Yang et al. “DAGAN: deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction”. In: *IEEE transactions on medical imaging* 37.6 (2017), pp. 1310–1321.

[48] Jun-Yan Zhu et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks*. 2020. arXiv: 1703.10593 [cs.CV].

[49] Francesco Tonolini et al. “Variational Inference for Computational Imaging Inverse Problems”. In: *Journal of Machine Learning Research* 21.179 (2020), pp. 1–46. URL: http://jmlr.org/papers/v21/20-151.html.

[50] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. “Deep Image Prior”. In: *International Journal of Computer Vision* 128.7 (Mar. 2017), pp. 1867–1888. ISSN: 1573-1405. DOI: 10.1007/s11263-020-01303-4. URL: http://dx.doi.org/10.1007/s11263-020-01303-4.

[51] Gary Mataev, Michael Elad, and Peyman Milanfar. *DeepRED: Deep Image Prior Powered by RED*. 2018. arXiv: 1903.10176 [cs.CV].

[52] Iksung Kang et al. “Simultaneous spectral recovery and CMOS micro-LED holography with an untrained deep neural network”. In: *Optica* 9.10 (2022), pp. 1149–1155.

[53] Iksung Kang et al. “Recurrent neural network reveals transparent objects through scattering media”. In: *Optics Express* 29.4 (2021), pp. 5316–5326.

[54] Yair Rivenson et al. “Phase recovery and holographic image reconstruction using deep learning in neural networks”. In: *Light: Science & Applications* 7.2 (Oct. 2017), pp. 17141–17141. DOI: 10.1038/lsa.2017.141. URL: https://doi.org/10.1038/lsa.2017.141.

[55] Ryoichi Horisaki, Ryosuke Takagi, and Jun Tanida. “Deep-learning-generated holography”. In: *Applied Optics* 57.14 (2018), pp. 3859–3863.

[56] Yair Rivenson, Yichen Wu, and Aydogan Ozcan. “Deep learning in holography and coherent imaging”. In: *Light: Science & Applications* 8.1 (2019), pp. 1–8.

[57] Tairan Liu et al. “Deep learning-based super-resolution in coherent imaging systems”. In: *Scientific reports* 9.1 (2019), p. 3926.

[58] M Hossein Eybposh et al. “DeepCGH: 3D computer-generated holography using deep learning”. In: *Optics Express* 28.18 (2020), pp. 26636–26650.

[59] Yunzhe Li, Yujia Xue, and Lei Tian. “Deep speckle correlation: a deep learning approach toward scalable imaging through scattering media”. In: *Optica* 5.10 (2018), pp. 1181–1190.

[60] Fangshu Yang et al. “Deep-learning projector for optical diffraction tomography”. In: *Optics Express* 28.3 (2020), pp. 3905–3921.

[61] Kevin C Zhou and Roarke Horstmeyer. "Diffraction tomography with a deep image prior". In: *Optics Express* 28.9 (2020), pp. 12872–12896.

[62] Thanh Nguyen et al. "Deep learning approach for Fourier ptychography microscopy". In: *Optics Express* 26.20 (2018), pp. 26470–26484.

[63] Ziqiao Guan and Esther H Tsai. *PtychoNet: Fast and high quality phase retrieval for ptychography*. Tech. rep. Brookhaven National Lab.(BNL), Upton, NY (United States), 2019.

[64] Yican Chen et al. "U-net CNN based Fourier ptychography". In: *arXiv preprint arXiv:2003.07460* (2020).

[65] Omri Wengrowicz et al. "Deep neural networks in single-shot ptychography". In: *Optics Express* 28.12 (2020), pp. 17511–17520.

[66] Ulugbek S Kamilov, Hassan Mansour, and Brendt Wohlberg. "A plug-and-play priors approach for solving nonlinear imaging inverse problems". In: *IEEE Signal Processing Letters* 24.12 (2017), pp. 1872–1876.

[67] Yair Rivenson et al. "Deep learning microscopy". In: *Optica* 4.11 (2017), pp. 1437–1443.

[68] Kevin de Haan et al. "Deep-learning-based image reconstruction and enhancement in optical microscopy". In: *Proceedings of the IEEE* 108.1 (2019), pp. 30–50.

[69] Yujia Xue et al. "Reliable deep-learning-based phase imaging with uncertainty quantification". In: *Optica* 6.5 (2019), pp. 618–629.

[70] Weijie Gan et al. "PtychoDV: Vision Transformer-Based Deep Unrolling Network for Ptychographic Image Reconstruction". In: *IEEE Open Journal of Signal Processing* (2024).

[71] Alexey Dosovitskiy et al. "An image is worth 16x16 words: Transformers for image recognition at scale". In: *arXiv preprint arXiv:2010.11929* (2020).

[72] Jo Schlemper et al. "A deep cascade of convolutional neural networks for dynamic MR image reconstruction". In: *IEEE transactions on Medical Imaging* 37.2 (2017), pp. 491–503.

[73] Kerstin Hammernik et al. "Learning a variational network for reconstruction of accelerated MRI data". In: *Magnetic resonance in medicine* 79.6 (2018), pp. 3055–3071.

[74] Yu Sun, Brendt Wohlberg, and Ulugbek S Kamilov. "An online plug-and-play algorithm for regularized image reconstruction". In: *IEEE Transactions on Computational Imaging* 5.3 (2019), pp. 395–408.

[75] Kyong Hwan Jin et al. "Deep convolutional neural network for inverse problems in imaging". In: *IEEE Transactions on Image Processing* 26.9 (2017), pp. 4509–4522.

[76] Maziar Raissi, Paris Perdikaris, and George E Karniadakis. "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations". In: *Journal of Computational physics* 378 (2019), pp. 686–707.

[77] George Em Karniadakis et al. “Physics-informed machine learning”. In: *Nature Reviews Physics* 3.6 (2021), pp. 422–440.

[78] Jia Deng et al. “ImageNet: A large-scale hierarchical image database”. In: *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 2009, pp. 248–255. DOI: 10.1109/CVPR.2009.5206848.

[79] Weiwen Wu et al. “Stabilizing deep tomographic reconstruction networks”. In: *arXiv:2008.01846* (2020).

[80] Li Deng. “The MNIST Database of Handwritten Digit Images for Machine Learning Research [Best of the Web]”. In: *IEEE Signal Processing Magazine* 29.6 (2012), pp. 141–142. DOI: 10.1109/MSP.2012.2211477.

[81] Johann Radon. “Uber die Bestimmung von Funktionen durch ihre Integralwerte langs gewissez Mannigfaltigheiten, Ber”. In: *Verh. Sachs. Akad. Wiss. Leipzig, Math Phys Klass* 69 (1917).

[82] Allan Macleod Cormack. “Representation of a function by its line integrals, with some radiological applications”. In: *Journal of applied physics* 34.9 (1963), pp. 2722–2727.

[83] Sigurdur Helgason and S Helgason. *The radon transform*. Vol. 2. Springer, 1999.

[84] Willi A Kalender. “X-ray computed tomography”. In: *Physics in medicine & Biology* 51.13 (2006), R29.

[85] Mark E. Davison. “The Ill-Conditioned Nature of the Limited Angle Tomography Problem”. In: *SIAM Journal on Applied Mathematics* 43.2 (1983), pp. 428–448. DOI: 10.1137/0143028. eprint: https://doi.org/10.1137/0143028. URL: https://doi.org/10.1137/0143028.

[86] Jürgen Frikel. “Reconstructions in limited angle X-ray tomography: Characterization of classical reconstructions and adapted curvelet sparse regularization”. PhD thesis. Technische Universität München, 2013.

[87] Takuso Sato et al. “Tomographic image reconstruction from limited projections using iterative revisions in image and transform spaces”. In: *Applied Optics* 20.3 (1981), pp. 395–399.

[88] Dean Verhoeven. “Limited-data computed tomography algorithms for the physical sciences”. In: *Applied Optics* 32.20 (1993), pp. 3736–3754.

[89] C. Bouman and K. Sauer. “A generalized Gaussian image model for edge-preserving MAP estimation”. In: *IEEE Transactions on Image Processing* 2 (1993), pp. 296–310.

[90] Daniil Kazantsev et al. “Employing temporal self-similarity across the entire time domain in computed tomography reconstruction”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 373.2043 (2015), p. 20140389.

[91] Aicha Allag et al. “X-rays tomographic reconstruction images using proximal methods based on L1 norm and TV regularization”. In: *Procedia Computer Science* 127 (2018), pp. 236–245.

[92] Wenkun Zhang et al. “Multi-energy CT reconstruction using tensor nonlocal similarity and spatial sparsity regularization”. In: *Quantitative Imaging in Medicine and Surgery* 10.10 (2020), p. 1940.

[93] Alexandre Goy et al. “High-resolution limited-angle phase tomography of dense layered objects using deep neural networks”. In: *Proceedings of the National Academy of Sciences* 116.40 (2019), pp. 19848–19856.

[94] Ji He, Yongbo Wang, and Jianhua Ma. “Radon inversion via deep learning”. In: *IEEE Transactions on Medical Imaging* 39.6 (2020), pp. 2076–2087.

[95] Z. H. Levine et al. “Scatter corrections in X-ray computed tomography: A physics-based analysis”. In: *J. Res. of the Natl. Inst. of Stand. and Tech.* 124 (2019), p. 124013.

[96] Z. H. Levine et al. “X-ray Computed Tomography using Partially Coherent Fresnel Diffraction with Application to an Optical Fiber”. In: *Opt. Express* 29 (2021), pp. 1788–1804.

[97] Ken Sauer and Charles Bouman. “A Local Update Strategy for Iterative Reconstruction from Projections”. In: *IEEE Transactions on Signal Processing* 41.2 (1993), pp. 534–548.

[98] Roger Fletcher. *Practical methods of optimization*. John Wiley & Sons, 2013.

[99] Mehdi Mirza and Simon Osindero. “Conditional generative adversarial nets”. In: *arXiv:1411.1784* (2014).

[100] Ian Goodfellow et al. “Generative adversarial nets”. In: *Advances in Neural Information Processing Systems* 27 (2014), pp. 1–9.

[101] Takeru Miyato et al. “Spectral normalization for generative adversarial networks”. In: *arXiv:1802.05957* (2018).

[102] Zhen Guo. *Physics-assisted Generative Adversarial Network for X-Ray Tomography*. https://github.com/zguo0525/Physics-assisted-Generative-Adversarial-Network-for-X-Ray-Tomography. 2021.

[103] Lei Jimmy Ba and Rich Caruana. “Do deep nets really need to be deep?” In: *arXiv:1312.6184* (2013).

[104] Huiyu Wang et al. “Axial-DeepLab: Stand-alone axial-attention for panoptic segmentation”. In: *European Conference on Computer Vision*. Springer. 2020, pp. 108–126.

[105] Stéphane Mallat. “Group invariant scattering”. In: *Communications on Pure and Applied Mathematics* 65.10 (2012), pp. 1331–1398.

[106] Joakim Andén and Stéphane Mallat. “Deep scattering spectrum”. In: *IEEE Transactions on Signal Processing* 62.16 (2014), pp. 4114–4128.

[107] Jean-Benoit Delbrouck and Stéphane Dupont. “Modulating and attending the source image during encoding improves multimodal translation”. In: *arXiv:1712.03449* (2017).

[108] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. "U-Net: Convolutional networks for biomedical image segmentation". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2015, pp. 234–241.

[109] Kaiming He et al. "Deep residual learning for image recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016, pp. 770–778.

[110] Mathieu Andreux et al. "Kymatio: Scattering Transforms in Python". In: *Journal of Machine Learning Research* 21.60 (2020), pp. 1–6. URL: http://jmlr.org/papers/v21/19-047.html.

[111] Yanghao Li et al. "Adaptive batch normalization for practical domain adaptation". In: *Pattern Recognition* 80 (2018), pp. 109–117.

[112] Albert Reuther et al. "Interactive supercomputing on 40,000 cores for machine learning and data analysis". In: *2018 IEEE High Performance Extreme Computing Conference (HPEC)*. IEEE. 2018, pp. 1–6.

[113] Diederik P Kingma and Jimmy Ba. "Adam: A method for stochastic optimization". In: *arXiv:1412.6980* (2014).

[114] Martin Heusel et al. "GANs trained by a two time-scale update rule converge to a local Nash equilibrium". In: *Advances in Neural Information Processing Systems* 30 (2017).

[115] Jae Hyun Lim and Jong Chul Ye. "Geometric GAN". In: *arXiv preprint arXiv:1705.02894* (2017).

[116] Alec Radford, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks". In: *:1511.06434* (2015).

[117] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein generative adversarial networks". In: *International Conference on Machine Learning*. PMLR. 2017, pp. 214–223.

[118] Ishaan Gulrajani et al. "Improved training of Wasserstein GANs". In: *arXiv:1704.00028* (2017).

[119] Andrew Brock, Jeff Donahue, and Karen Simonyan. "Large scale GAN training for high fidelity natural image synthesis". In: *arXiv:1809.11096* (2018).

[120] Zinan Lin, Vyas Sekar, and Giulia Fanti. "Why Spectral Normalization Stabilizes GANs: Analysis and Improvements". In: *Thirty-Fifth Conference on Neural Information Processing Systems*. 2021.

[121] Haoyu Lan et al. "SC-GAN: 3D self-attention conditional GAN with spectral normalization for multi-modal neuroimaging synthesis". In: *bioRxiv* (2020). DOI: 10.1101/2020.06.09.143297. eprint: https://www.biorxiv.org/content/early/2020/06/11/2020.06.09.143297.full.pdf. URL: https://www.biorxiv.org/content/early/2020/06/11/2020.06.09.143297.

[122] Ilya Kavalerov, Wojciech Czaja, and Rama Chellappa. "A multi-class hinge loss for conditional GANs". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2021, pp. 1290–1299.

[123] David L Donoho and Michael Elad. "Optimally sparse representation in general (nonorthogonal) dictionaries via $\ell 1$ minimization". In: *Proceedings of the National Academy of Sciences* 100.5 (2003), pp. 2197–2202.

[124] Karol Gregor and Yann LeCun. "Learning fast approximations of sparse coding". In: *Proceedings of the 27th international conference on machine learning*. 2010, pp. 399–406.

[125] Michael Elad and Michal Aharon. "Image denoising via sparse and redundant representations over learned dictionaries". In: *IEEE Transactions on Image Processing* 15.12 (2006), pp. 3736–3745.

[126] Farah Deeba et al. "Sparse representation based computed tomography images reconstruction by coupled dictionary learning algorithm". In: *IET image Processing* 14.11 (2020), pp. 2365–2375.

[127] Jevgenija Rudzusika, Thomas Koehler, and Ozan Oktem. "Deep Learning-Based Dictionary Learning and Tomographic Image Reconstruction". In: *SIAM Journal on Imaging Sciences* 15.4 (2022), pp. 1729–1764.

[128] Tobias Würfl et al. "Deep learning computed tomography". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2016, pp. 432–440.

[129] Wei Wang et al. "An end-to-end deep network for reconstructing CT images directly from sparse sinograms". In: *IEEE Transactions on Computational Imaging* 6 (2020), pp. 1548–1560.

[130] Michael M Lell and Marc Kachelrieß. "Recent and upcoming technological developments in computed tomography: high speed, low dose, deep learning, multienergy". In: *Investigative Radiology* 55.1 (2020), pp. 8–19.

[131] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*. 2015. arXiv: 1505.04597 [cs.CV].

[132] Masashi Sugiyama and Motoaki Kawanabe. *Machine Learning in Non-Stationary Environments: Introduction to Covariate Shift Adaptation*. MIT Press, 2012.

[133] Kun Zhang et al. "Domain adaptation under target and conditional shift". In: *International Conference on Machine Learning*. PMLR. 2013, pp. 819–827.

[134] Jiashuo Liu et al. "Towards out-of-distribution generalization: A survey". In: *arXiv preprint arXiv:2108.13624* (2021).

[135] Yoseob Han and Jong Chul Ye. "Framing U-Net via deep convolutional framelets: Application to sparse-view CT". In: *IEEE transactions on medical imaging* 37.6 (2018), pp. 1418–1429.

[136] Doga Gunduzalp et al. *3D U-NetR: Low Dose Computed Tomography Reconstruction via Deep Learning and 3 Dimensional Convolutions*. 2022. arXiv: 2105.14130 [cs.CV].

[137] Karl Weiss, Taghi M Khoshgoftaar, and DingDing Wang. "A survey of transfer learning". In: *Journal of Big Data* 3.1 (2016), pp. 1–40.

[138] Zhen Guo. *Noise resilience deep reconstruction for X-ray tomography.* https://github.com/zguo0525/Noise-resilience-deep-reconstruction-for-X-ray-Tomography. Version 1.0.0. Nov. 2022.

[139] Patrick Schober, Christa Boer, and Lothar A Schwarte. "Correlation coefficients: appropriate use and interpretation". In: *Anesthesia & Analgesia* 126.5 (2018), pp. 1763–1768.

[140] K Yen, Eugene K Yen, and Roger G Johnston. "The ineffectiveness of the correlation coefficient for image comparisons". In: *Los Alamos National Laboratory report LA-UR-96-2474* (1996).

[141] Joan Bruna and Stéphane Mallat. "Invariant scattering convolution networks". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35.8 (2013), pp. 1872–1886.

[142] Mathieu Andreux et al. "Kymatio: Scattering Transforms in Python". In: *Journal of Machine Learning Research* 21.60 (2020), pp. 1–6. URL: http://jmlr.org/papers/v21/19-047.html.

[143] Alexandre Bernardino and José Santos-Victor. "A real-time gabor primal sketch for visual attention". In: *Iberian Conference on Pattern Recognition and Image Analysis*. Springer. 2005, pp. 335–342.

[144] Ignace Loris et al. "Tomographic inversion using l1-norm regularization of wavelet coefficients". In: *Geophysical Journal International* 170.1 (2007), pp. 359–370.

[145] Anja Borsdorf et al. "Wavelet based noise reduction in CT-images using correlation analysis". In: *IEEE transactions on medical imaging* 27.12 (2008), pp. 1685–1703.

[146] Harbir Antil, Zichao Wendy Di, and Ratna Khatri. "Bilevel optimization, deep learning and fractional Laplacian regularization with applications in tomography". In: *Inverse Problems* 36.6 (2020), p. 064001.

[147] Timothy F Scott et al. "Two-color single-photon photoinitiation and photoinhibition for subdiffraction photolithography". In: *Science* 324.5929 (2009), pp. 913–917.

[148] Yaoyu Cao et al. "High-photosensitive resin for super-resolution direct-laser-writing based on photoinhibited polymerization". In: *Optics Express* 19.20 (2011), pp. 19486–19494.

[149] Martin P De Beer et al. "Rapid, continuous additive manufacturing by volumetric polymerization inhibition patterning". In: *Science advances* 5.1 (2019), eaau8723.

[150] Xing Lin et al. "All-optical machine learning using diffractive deep neural networks". In: *Science* 361.6406 (2018), pp. 1004–1008.

[151] Tao Yan et al. "Fourier-space diffractive deep neural network". In: *Physical review letters* 123.2 (2019), p. 023901.

[152] Deniz Mengu et al. “At the intersection of optics and deep learning: statistical inference, computing, and inverse design”. In: *Advances in Optics and Photonics* 14.2 (2022), pp. 209–290.

[153] Henry Arguello et al. “Deep optical coding design in computational imaging: a data-driven framework”. In: *IEEE Signal Processing Magazine* 40.2 (2023), pp. 75–88.

[154] JR Lawrence, FT O’Neill, and JT Sheridan. “Photopolymer holographic recording material”. In: *Optik* 112.10 (2001), pp. 449–463.

[155] Michael R Gleeson and John T Sheridan. “A review of the modelling of free-radical photopolymerization in the formation of holographic gratings”. In: *Journal of optics A: pure and applied optics* 11.2 (2009), p. 024008.

[156] Michael R Gleeson and John T Sheridan. “Nonlocal photopolymerization kinetics including multiple termination mechanisms and dark reactions. Part I. Modeling”. In: *JOSA B* 26.9 (2009), pp. 1736–1745.

[157] Klaus Cicha et al. “Evaluation of 3D structures fabricated with two-photon-photopolymerization by using FTIR spectroscopy”. In: *Journal of Applied Physics* 110.6 (2011).

[158] Christopher N LaFratta and Tommaso Baldacchini. “Two-photon polymerization metrology: Characterization methods of mechanisms and microstructures”. In: *Micromachines* 8.4 (2017), p. 101.

[159] Qi Ge et al. “Projection micro stereolithography based 3D printing and its applications”. In: *International Journal of Extreme Manufacturing* 2.2 (2020), p. 022004.

[160] B-J De Gans, Paul C Duineveld, and Ulrich S Schubert. “Inkjet printing of polymers: state of the art and future developments”. In: *Advanced materials* 16.3 (2004), pp. 203–213.

[161] Madhusudan Singh et al. “Inkjet printing process and its applications”. In: *Advanced materials* 22.6 (2010), pp. 673–685.

[162] Jill Z Manapat et al. “3D printing of polymer nanocomposites via stereolithography”. In: *Macromolecular Materials and Engineering* 302.9 (2017), p. 1600553.

[163] Kwang-Sup Lee et al. “Recent developments in the use of two-photon polymerization in precise 2D and 3D microfabrications”. In: *Polymers for advanced technologies* 17.2 (2006), pp. 72–82.

[164] Xiaoqin Zhou, Yihong Hou, and Jieqiong Lin. “A review on the processing accuracy of two-photon polymerization”. In: *Aip Advances* 5.3 (2015).

[165] Tommaso Baldacchini. *Three-dimensional microfabrication using two-photon polymerization: fundamentals, technology, and applications*. William Andrew, 2015.

[166] John V Kelly et al. “Temporal analysis of grating formation in photopolymer using the nonlocal polymerization-driven diffusion model”. In: *Optics Express* 13.18 (2005), pp. 6990–7004.

[167] Suyong Kim et al. “Stiff neural ordinary differential equations”. In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 31.9 (Sept. 2021). ISSN: 1089-7682. DOI: 10.1063/5.0060697. URL: http://dx.doi.org/10.1063/5.0060697.

[168] M. Calvo, J.I. Montijano, and L. Randez. “A fifth-order interpolant for the Dormand and Prince Runge-Kutta method”. In: *Journal of Computational and Applied Mathematics* 29.1 (1990), pp. 91–100. ISSN: 0377-0427. DOI: https://doi.org/10.1016/0377-0427(90)90198-9. URL: https://www.sciencedirect.com/science/article/pii/0377042790901989.

[169] Wolfgang Arendt and Mahamadi Warma. “The Laplacian with Robin boundary conditions on arbitrary domains”. In: *Potential Analysis* 19 (2003), pp. 341–363.

[170] Henry N Chapman and Keith A Nugent. “Coherent lensless X-ray imaging”. In: *Nature Photonics* 4.12 (2010), pp. 833–839.

[171] Michael W Davidson and Mortimer Abramowitz. “Optical microscopy”. In: *Encyclopedia of imaging science and technology* 2.1106-1141 (2002), p. 120.

[172] Jerome Mertz. *Introduction to optical microscopy*. Cambridge University Press, 2019.

[173] Eikichi Yamashita. “Analysis methods for electromagnetic wave problems”. In: (1990).

[174] Ulugbek S Kamilov et al. “Optical tomographic image reconstruction based on beam propagation and sparse regularization”. In: *IEEE Transactions on Computational Imaging* 2.1 (2016), pp. 59–70.