2019

Computational Imaging with Limited Photon Budget

Zheyuan Zhu
University of Central Florida

COMPUTATIONAL IMAGING WITH LIMITED PHOTON BUDGET

By

ZHEYUAN ZHU
B.S. Nanjing University, 2015
M.S. University of Central Florida, 2020

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in CREOL, The College of Optics and Photonics
at the University of Central Florida
Orlando, Florida

Summer Term
2020

Major Professor: Sean Pang

ABSTRACT

The capability of retrieving the image/signal of interest from extremely low photon flux is attractive in scientific, industrial, and medical imaging applications. Conventional imaging modalities and reconstruction algorithms rely on hundreds to thousands of photons per pixel (or per measurement) to ensure enough signal-to-noise (SNR) ratio for extracting the image/signal of interest. Unfortunately, the potential of radiation or photon damage prohibits high SNR measurements in dose-sensitive diagnosis scenarios. In addition, imaging systems utilizing inherently weak signals as contrast mechanism, such as X-ray scattering-based tomography, or attosecond pulse retrieval from the streaking trace, entail prolonged integration time to acquire hundreds of photons, thus rendering high SNR measurement impractical.

This dissertation addresses the problem of imaging from limited photon budget when high SNR measurements are either prohibitive or impractical. A statistical image reconstruction framework based on the knowledge of the image-formation process and the noise model of the measurement system has been constructed and successfully demonstrated on two imaging platforms – photon-counting X-ray imaging, and attosecond pulse retrieval. For photon-counting X-ray imaging, the statistical image reconstruction framework achieves high-fidelity X-ray projection and tomographic image reconstruction from as low as 16 photons per pixel on average. The capability of our framework in modeling the reconstruction error opens the opportunity of designing the optimal strategies to distribute a fixed photon budget for region-of-interest (ROI) reconstruction,

paving the way for radiation dose management in an imaging-specific task. For attosecond pulse retrieval, a learning-based framework has been incorporated into the statistical image reconstruction to retrieve the attosecond pulses from the noisy streaking traces. Quantitative study on the required signal-to-noise ratio for satisfactory pulse retrieval enabled by our framework provides a guideline to future attosecond streaking experiments. In addition, resolving the ambiguities in the streaking process due to the carrier envelop phase has also been demonstrated with our statistical reconstruction framework.

ACKNOWLEDGEMENTS

I would like to acknowledge the support from everyone who helped me during my career as a doctorate student. I would like to thank my supervisor Dr. Shuo Pang for his invaluable guidance and support on my research, as well as offering me plenty of exciting projects for me to explore and expand my knowledge.

I gratefully acknowledge our collaborators, Dr. Alexander Katsevich, Dr. Hsin-Hsiung Huang, Dr. Zenghu Chang, and Dr. Guifang Li. Working together with you opens my perspective toward interdisciplinary fields across optics and photonics, applied mathematics, and physics.

I would also like to express my thanks to my current and past colleagues, Yangyang Sun, Ryan Ellis, Joseph Ulseth, in Optical Imaging System Lab. The discussions with you over the past years on experiment challenges, programming tricks, project ideas etc., have been fruitful and inspiring. Thanks for your continuous support and encouragement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Imaging from low photon budget is attractive in many photon-starving imaging systems [1]. Photo-starving scenarios can occur when the image is formed from intrinsically weak signals. This scenario includes X-ray scattering tomography, including Rayleigh scattering or Compton scattering whose signal, despite providing enhanced material, is 10^2~10^3 times weaker than the transmitted beam [2]. The same situation also happens in attosecond pulse retrieval from the streaking trace [3], in which the photon flux of the attosecond pulse used in the streaking experiment is orders of magnitude weaker than the femtosecond pump laser [4]. These scenarios typically require hours long imaging time to accumulate enough photons for satisfactory signal-to-noise ratio on the detector, which can be impractical in real-time imaging applications. Another photon-starving scenario occurs when photon damage of the specimen prohibits the use of large photon flux in the measurements, and thus reducing the signal-to-noise ratio (SNR) of the detected signal. This scenario can be found in virtually all X-ray imaging systems, including radiography (X-ray projection image), computed tomography (CT), and spectroscopy, because the excessive ionizing radiation can induce preferment damage, which manifests either as conformational changes in protein for biological samples [5,6], or functional changes in the microstructure of processors, memory chips, and other integrated circuits [7,8] in industrial parts inspection.

Recently, ingenious measurement system designs have been proposed to partially mitigate the photon-starving scenarios. In X-ray imaging, solutions such as parallelizing the data acquisition [9,10], or reducing the data acquisition only to the region-of-interest (ROI) [2,11] have demonstrated the potential to reduce the imaging time by one order of magnitude. Despite these attempts to accelerate the imaging process, the amount of radiation per area or volume remain the same, if the same SNR as a point-by-point imaging system on the reconstructed image were to be achieved [9]. Moreover, these unconventional data acquisition schemes necessities careful alignment and calibration, intricating the overall operation of these imaging systems compared to their point-by-point scanning counterpart. Therefore, in most photon-starving scenarios that demand easy-to-use setup and low radiation exposure, high SNR measurements are still not readily feasible.

Since increasing the signal level cannot be easily accomplished with a simple modification to the imaging system, reducing the noise in the measurement becomes a natural choice. Detectors that eliminate as much thermal noise as possible via improved operation environment and integrated pulse counting, sorting, and shape-specific gating circuits can be used. These detectors enable the measurement of the signal at the precision of single photons in both visible and X-ray imaging [12–14]. Nevertheless, the received photon counts still fluctuate due to the discrete nature of photons, a phenomenon termed “photon-limited noise”, even the measurement can be completely deprived of thermal noise. Up till recently, computational imaging reconstruction has not yet incorporated the complete physical process of the noise and image formation to cope with the systems’

hardware properties. Most of the conventional image reconstruction algorithms can be summarized as two alternating steps that enforces the measurement consistency and signal/image constraints in an iterative process [15–17]. These algorithms are designed to minimize the mean squared error (MSE) between the measured signal and simulated measurement from the estimated signal/image. The fundamental design principle behind these iterative algorithms implies an equivalent of Gaussian (thermal) noise on the detector side, and thus would fail when handling photon-counting signals with low counts due to the discrepancy in their underlying assumptions on the physical and noise model.

In this dissertation, we introduce a statistical image reconstruction framework for solving a variety of photon-starving imaging problems, including X-ray transmission and tomographic imaging, region-of-interest tomography, and attosecond pulse retrieval. The proposal is organized as follows. Chapter 2 discusses the physical process of photon-counting measurements; and proposes a statistical image/signal inference framework based on the estimation of the posterior probability of possible images/signals given a photon-counting measurement. Learning-based posterior estimation is then introduced to obviate the iterative inference process and accelerate the image/signal retrieval. In chapter 3 and 4, we present several examples of statistical image retrieval, including image reconstruction from photon-counting X-ray projection and CT measurements, as well as pulse retrieval from attosecond streaking traces with limited photon flux. For photon-counting CT, we will apply our statistical reconstruction framework to study the optimal photon budget allocation for region-of-interest reconstruction in CT. For attosecond pulse retrieval, the proposed statistical reconstruction framework is combined

with a learning-based method to model the pulse reconstruction error and uncertainty at various noise levels, which can serve as guidelines to future attosecond streaking experiments. Because the proposed framework can capture the distribution of all possible signals given an observation, we will also demonstrate its capability in resolving phase-shifting ambiguity in the streaking process. Chapter 5 concludes the findings and results throughout this study.

CHPATER 2: PHOTON-COUNTING IMAGE FORMATION AND RECONSTRUCTION

In this chapter we establish the statistical theory behind photon-counting detection, derive the statistical estimation of the signal/image of interest from the photon-counting measurements, and show its asymptotic behavior as the photon counts increases. This chapter focuses on the discussion of photon-counting in X-ray imaging, where attenuation is the primary contrast mechanism. At the end of this chapter, we will extend the discussion to a general imaging model.

2.1 Statistical photon counting model

Figure 1 illustrates the readout from a detector operating in the ideal photon-counting mode, in which all the thermal noise has been filtered out via pulse censoring circuitry. Avalanche photodiode or X-ray photon-counting detectors [12–14] are a few examples that can be modeled with this ideal photon-counting detection scheme. Consider a pencil-beam X-ray source and a photon-counting detector. Assume the counting rate $1/\Delta t$ (interval Δt) is much faster (shorter) than the rate (interval) of photon arrival instances, such that no two photon instances are registered within the same counting interval. The probability of detecting a photon instance within one interval when no sample is present between the source and the detector is denoted as λ. In experiment, λ needs to be calibrated beforehand, and is the proportional to the product between source flux and detector quantum efficiency.

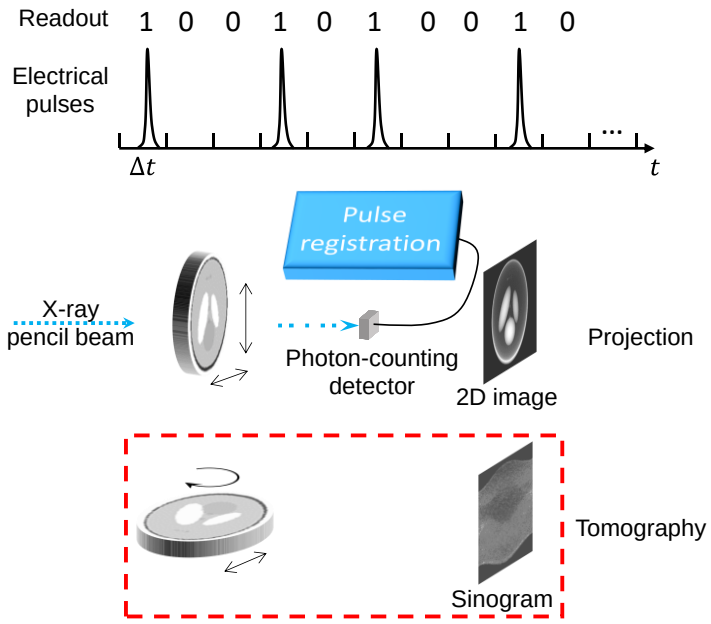

Figure 1: Illustration of the photon-counting scheme for X-ray imaging, including projection and tomography setups.

When an object, discretized by $\mathbf{f}$, is placed into the pencil beam, the probability of receiving one photon in each time interval is reduced exponentially

$$T_j = \lambda \exp\left(-\sum_{i=1}^{n} \mathbf{A}_{ij}\mathbf{f}_i\right) \tag{1}$$

where the subscript $i = 1,2,\ldots,n$ represents the index of the discretized object attenuation map $\mathbf{f}$; and $j = 1,2 \ldots m$ represents the index of discretized pencil-beam measurements. The matrix $\mathbf{A}$ establishes the linear relation between the object and the measurement. For X-ray projection imaging, $\mathbf{A}$ is the identity matrix, and for tomography, $\mathbf{A}$ represents the Radon transform matrix constructed from the distance-driven ray-tracing model [18]. The joint probability of detecting the $\mathbf{r}$-th photon at $\mathbf{g}$-th

time interval follows the negative binomial distribution $\mathbf{g}\sim NB(\mathbf{r},\mathbf{T})$, ($\mathbf{r}\in\mathbb{N}_{+}^{m}$,$\mathbf{g}\in\mathbb{N}_{+}^{m}$), whose probability mass function (PMF) is

$$p(\mathbf{g}|\mathbf{f};\mathbf{r})=\prod_{j=1}^{m}\binom{g_j-1}{r_j-1}\left(1-T_j\right)^{g_j-r_j}{T_j}^{r_j}, \tag{2}$$

where $\mathbf{g}=(g_1,g_2,\dots g_m)$ is the total number of time intervals that has elapsed upon the arrival of the first $\mathbf{r}=(r_1,r_2,\dots r_m)$ photons at each pencil beam, j. If instead the number of photons are counted within a predefined period $\mathbf{g}\Delta t$, the joint probability of receiving $\mathbf{r}$ photons for each pencil beam follows binomial distribution $\mathbf{r}\sim B(\mathbf{g},\mathbf{T})$ ($\mathbf{g}\in\mathbb{N}_{+}^{m},\mathbf{r}\in\mathbb{N}^{m}$)

$$p(\mathbf{r}|\mathbf{f};\mathbf{g})=\prod_{j=1}^{m}\binom{g_j}{r_j}\left(1-T_j\right)^{g_j-r_j}{T_j}^{r_j}. \tag{3}$$

Notice that if the transmitted photon flux T_j is small and the number of intervals g_j is sufficiently large, Equation (3) can be approximated by $\mathbf{r}\sim Poisson(\mathbf{gT})$ where the detected photons within the time period $\mathbf{g}\Delta t$ follows Poisson distribution

$$p(\mathbf{r}|\mathbf{f};\mathbf{g})=\prod_{j=1}^{m}\frac{{T_j}^{r_j}\exp(-g_jT_j)}{r_j!}. \tag{4}$$

2.2 Bayesian inference for image/signal reconstruction

From a statistical perspective, image reconstruction is the process of estimating signal/image of interest $\hat{\mathbf{f}}$, from the measurements $\mathbf{g}$, knowing the combined noise model and image formation process $p(\mathbf{g}|\mathbf{f};\mathbf{r})$. Here we consider the binomial noise

model (Equation (3)), in which $\mathbf{r}$ represents the number of photons to accumulate for each pencil beam, and is a set of parameters of the imaging system.

The reconstructed image/signal of interest $\hat{\mathbf{f}}$ should be the one that minimizes the negative posterior likelihood $p(\mathbf{f}|\mathbf{g};\mathbf{r})$ (or its logarithm $L(\mathbf{f}|\mathbf{g};\mathbf{r})$) [19]. This statistical estimation scheme is termed maximum-a-posteriori, MAP. Applying Bayes' rule in the calculation of posterior likelihood, the estimator can be formulated as

$$\hat{\mathbf{f}} = \underset{\mathbf{f}}{\operatorname{argmin}}\, L(\mathbf{f}|\mathbf{g};\mathbf{r}) = \underset{\mathbf{f}}{\operatorname{argmin}} -\log\left[\frac{p(\mathbf{g}|\mathbf{f};\mathbf{r})\pi(\mathbf{f})}{p(\mathbf{g};\mathbf{r})}\right] \tag{5}$$

where $\pi(\mathbf{f})$ is the prior distribution covering all plausible images/signals, $\mathbf{f}$, and its logarithm enforces constraint or regularization on the reconstructed image/signal. The marginal distribution of the measurement $\log p(\mathbf{g};\mathbf{r})$ is independent of $\mathbf{f}$, and thus not included in the optimization. Let us assume $\pi(\mathbf{f})$ follows the distribution

$$\pi(\mathbf{f}) = C_{\beta,\mathbf{D}} \exp\left(-\tau\|\mathbf{D}\mathbf{f}\|_p^2\right), \tag{6}$$

where $C_{\beta,\mathbf{D}}$ is the normalization factor; $\mathbf{D}$ projects $\mathbf{f}$ onto the domain $\mathbf{u} = \mathbf{D}\mathbf{f}$ in which the image/signal is sparse, such as total-variation (TV) [20] or wavelet [21]; $\|\cdot\|_p$ denotes the L-p norm of the vector; τ adjusts the variance of the prior distribution, which in turn controls the strength of the regularization.

For image reconstruction from photon-counting measurements with a pre-defined photon number to accumulate at each pencil beam, the negative log-posterior distribution $L(\mathbf{f}|\mathbf{g};\mathbf{r})$ of waiting $\mathbf{g}$ intervals for $\mathbf{r}$ photons given the sample prior $\pi(\mathbf{f})$ is

$$L(\mathbf{f}|\mathbf{g};\mathbf{r}) = -\log\left[\frac{p(\mathbf{g}|\mathbf{f};\mathbf{r})\pi(\mathbf{f})}{p(\mathbf{g};\mathbf{r})}\right]$$

$$= -\sum_{j=1}^{m}\left\{\log\binom{g_j-1}{r_j-1} + (g_j - r_j)\log\left[1-\lambda\exp\left(-\sum_{i=1}^{n}A_{ji}f_i\right)\right] - r_j\sum_{i=1}^{n}A_{ji}f_i\right\} \quad (7)$$

$$+ \|\mathbf{Df}\|_p^2 + \log p(\mathbf{g};\mathbf{r}).$$

The minimization of Equation (7) is solved numerically using a gradient-descent method, such as a modified SPIRAL-TAP [16], which is a two-step, gradient-descent algorithm combined with regularization enforcement in each iteration. The gradient and Hessian of the negative binomial log-likelihood $-\log p(\mathbf{g}|\mathbf{f};\mathbf{r})$ are respectively

$$\nabla_{\mathbf{f}}(-\log p(\mathbf{g}|\mathbf{f};\mathbf{r})) = \mathbf{A}^T\left(\mathbf{r} - \frac{\lambda(\mathbf{g}-\mathbf{r})\odot\exp(-\mathbf{Af})}{1-\lambda\exp(-\mathbf{Af})}\right), \quad (8)$$

$$\mathbf{H}(-\log p(\mathbf{g}|\mathbf{f};\mathbf{r})) = \mathbf{A}^T diag\left(\frac{\lambda(\mathbf{g}-\mathbf{r})\odot\exp(-\mathbf{Af})}{(1-\lambda\exp(-\mathbf{Af}))^2}\right)\mathbf{A}. \quad (9)$$

Because the probability λ of detecting one photon within each time interval is on the order of 10^{-2}, it is very rare ($<10^{-4}$) to detect multiple photons in one interval. As a result, $g_j - r_j$ has a high probability of being non-negative, suggesting that the Hessian has a high probability of being positive semidefinite for non-negative object $\mathbf{f} \geq \mathbf{0}$. This ensures high probability of $L(\mathbf{f}|\mathbf{g};\mathbf{r})$ being a convex function with respect to $\mathbf{f}$, thus guarantees the convergence of the algorithm if the positivity of $\mathbf{f}$ is imposed throughout the iterations. The step of the gradient-descent is chosen according to the modified Barzilai-Borwein method described in Ref. [16], and is inversely proportional to $\Delta\mathbf{f}^{\boldsymbol{T}}\mathbf{H}\Delta\mathbf{f}$, where $\Delta\mathbf{f}$ is the update to the solution between adjacent iterations. Notice that Equation

(9) contains a singular point on the denominator. A maximum step size was enforced to avoid moving the solution across this singular point.

2.3 Learning-based posterior estimation

The essence of statistical image reconstruction is to write down the posterior distribution of all possible images/signals given the measurement $p(\mathbf{f}|\mathbf{g};\mathbf{r})$, from which we can then pick the image/signal of interest that maximizes it. A high-fidelity image reconstruction entails accurately modeling the image formation and noise process, as well as providing a prior distribution that can discriminate all plausible images against non-realistic ones. Based on Bayesian framework (Equation (5)), using a generative model to derive a more accurate prior distribution than Equation(6) becomes a logical follow-up [23–25]. The prior $\pi(\mathbf{f})$ can be trained on a series of observations $\{\mathbf{f}_i\}_{i=1}^{N}$. Though promising retrieval results has been demonstrated, the optimization remains a lengthy iterative process. In this section, we will describe a posterior estimation method based on conditional variational generative network, CVGN for short, to directly capture the posterior distribution $p(\mathbf{f}|\mathbf{g};\mathbf{r})$, for solving image/signal reconstruction problems. The proposed learning-based framework is a single-pass inference process through the trained neural networks, capable of handling the ambiguity in the image-formation process, as well as capturing the uncertainty of reconstructed image given a noisy measurement. In the following discussions, the parameters of imaging system $\mathbf{r}$ is treated as a constant and omitted from the probability distributions for simplification of the notation.

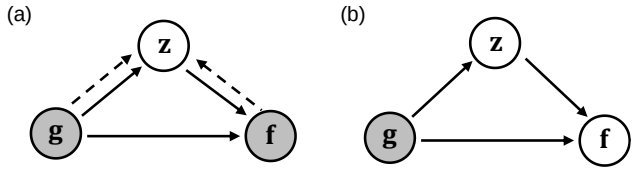

Figure 2: Directed graphical model (solid lines) of our proposed image/signal retrieval network, which contains an inference model (a) and a retrieval model (b). The image/signal retrieval process is parameterized by θ. Training of the parameters θ is assisted by introducing variational inference process $\boldsymbol{q_\phi}(\mathbf{z}|\mathbf{f},\mathbf{g})$ (dashed lines). Each arrow represents a conditional probability. Variables in gray contain observable data in their respective models.

The learning-based image reconstruction framework models the distribution of all possible images via a mixture of Gaussian distributions and parameterizing it with a latent variable $\mathbf{z}$,

$$p_\theta(\mathbf{f}|\mathbf{g}) = \int p_\theta(\mathbf{f}|\mathbf{z},\mathbf{g})p_\theta(\mathbf{z}|\mathbf{g})d\mathbf{z}. \tag{10}$$

During the image reconstruction process, the latent variable $\mathbf{z}$ was sampled from the conditional prior $p_\theta(\mathbf{z}|\mathbf{g})$ given measurement $\mathbf{g}$, and the retrieved signal $\mathbf{f}$ is generated from the conditional variational distribution $p_\theta(\mathbf{f}|\mathbf{z},\mathbf{g})$. Both $p_\theta(\mathbf{z}|\mathbf{g})$ and $p_\theta(\mathbf{f}|\mathbf{z},\mathbf{g})$ are assumed to be high-dimensional Gaussian distributions with diagonal covariance matrix, $p_\theta(\mathbf{z}|\mathbf{g}) = \mathcal{N}(\mathbf{z};\boldsymbol{\mu}_\mathbf{z}^{(\theta)}(\mathbf{g}), diag\left(\left[\boldsymbol{\sigma}_\mathbf{z}^{(\theta)}(\mathbf{g})\right]^2\right))$, $p_\theta(\mathbf{f}|\mathbf{z},\mathbf{g}) = \mathcal{N}(\mathbf{f};\boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{g},\mathbf{z}),\beta\mathbf{I})$, where the mean and variance parameters$\boldsymbol{\mu}_\mathbf{z}^{(\theta)}(\mathbf{g})$,$\boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{g},\mathbf{z})$ and $\boldsymbol{\sigma}_\mathbf{z}^{(\gamma)}(\mathbf{g})$ are implemented by neural networks with parameter θ [22], and β is a hyper parameter that determines the covariance of the posterior distribution.

The training process of conditional variational generative network maximizes the joint log-likelihood $\log p_\theta(\mathbf{f}|\mathbf{g}) = \sum_{i=1}^{N} \log p_\theta(\mathbf{f}_i|\mathbf{g}_i)$ of observing the pulse parameters vs trace pairs $\{(\mathbf{f}_i,\mathbf{g}_i), i = 1,\dots,N\}$ in the dataset. Due to the intractable integral in Equation(10), a

lower bound of the log-likelihood is used as the objective function, $\mathcal{L}$, with the introduction of a recognition distribution $q_\phi(\mathbf{z}|\mathbf{f},\mathbf{g})$ [23]

$$\log p_\theta(\mathbf{f}_i|\mathbf{g}_i) \geq -KL\left(q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)||p_\theta(\mathbf{z}|\mathbf{g}_i)\right) + E_{q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)}(\log p_\theta(\mathbf{f}_i|\mathbf{z},\mathbf{g}_i)) \coloneqq \mathcal{L}, \tag{11}$$

where $q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$ captures the latent distribution conditioned on both the streaking trace and pulse parameters. If we model $q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$ as a multivariate Gaussian with diagonal covariance matrix $\mathcal{N}(\mathbf{z};\boldsymbol{\mu}_\mathbf{z}^{(\phi)}(\mathbf{f},\mathbf{g}), diag\left(\left[\boldsymbol{\sigma}_\mathbf{z}^{(\phi)}(\mathbf{f},\mathbf{g})\right]^2\right))$, whose mean and variance are also implemented by neural networks with parameter ϕ, the objective function to maximize has a closed form

$$\mathcal{L} = -\sum_{j=1}^{M}\left(\log\frac{\sigma_{ij}^{(\phi)}}{\sigma_{ij}^{(\theta)}} + \frac{\left(\mu_{ij}^{(\theta)}-\mu_{ij}^{(\phi)}\right)^2+\sigma_{ij}^{(\theta)^2}}{2\sigma_{ij}^{(\phi)^2}} - \frac{1}{2}\right) - \frac{1}{\beta L}\sum_{l=1}^{L}\left(\mathbf{f}_i - \boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{z}_l,\mathbf{g}_i)\right)^2, \tag{12}$$

where $\sigma_{ij}^{(\phi)}$ denotes the j-th index of the M-element vectors $\boldsymbol{\sigma}_\mathbf{z}^{(\phi)}(\mathbf{f}_i,\mathbf{g}_i)$; similar notations are applied to $\sigma_{ij}^{(\theta)}$, $\mu_{ij}^{(\phi)}$ and $\mu_{ij}^{(\theta)}$. The expectation in Equation (11) is approximated by sampling L instances of $\mathbf{z}$ from the distribution $q_\phi(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$ as $\{\mathbf{z}_l: l=1,\dots,L\}$.

2.4 Learning-based signal retrieval considering measurement consistency

Section 2.3 presents a learning-based method, termed CVGN, to approximate the posterior distribution of possible images/signals given a specific measurement. The

training phase of CVGN draws samples $\mathbf{z}$ from the recognition distribution $q_\phi(\mathbf{z}|\mathbf{f}_i, \mathbf{g}_i)$. During the reconstruction process, however, CVGN draws $\mathbf{z}$ from the conditional prior distribution$p_\theta(\mathbf{z}|\mathbf{g}_i)$. This inconsistency between the recognition distribution and conditional prior distribution was also recognized in Ref. [36]. When using the variational lower bound as the objective function, relying only on closing the KL-divergence between $q_\phi(\mathbf{z}|\mathbf{f}_i, \mathbf{g}_i)$ and $p_\theta(\mathbf{z}|\mathbf{g}_i)$ cannot provide effective training to the conditional prior. In the future, we plan to utilize the known measurement process in constructing a supplemental network to assist the training, as shown in Figure 3.

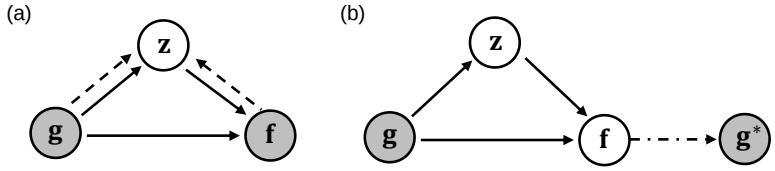

Figure 3: Proposed directed graphical models of both inference model (a) and retrieval model (b). The signal retrieval process is parameterized by θ. Training of the parameters θ is assisted by introducing (a) variational inference process $\boldsymbol{q_\phi}(\mathbf{z}|\mathbf{f}, \mathbf{g})$ (dashed lines), (b) the known physical model $\boldsymbol{A}(\cdot)$ of the measurement process (dot-dashed line). Each arrow represents a conditional probability. Variables in gray contain observable data in their respective models.

For the signal retrieval process, latent variable samples drawn from $p_\theta(\mathbf{z}|\mathbf{g}_i)$ capture the variance of all signals $\mathbf{f}$ that produce measurement $\mathbf{g}_i$. Naively replacing $q_\phi(\mathbf{z}|\mathbf{f}_i, \mathbf{g}_i)$ with $p_\theta(\mathbf{z}|\mathbf{g}_i)$ in the log-likelihood lower bound in Equation (11), in an attempt to keep $\mathbf{z}$ distribution consistency, amounts to comparing all possible signals $\mathbf{f} = \boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{z}_l, \mathbf{g}_i)$ given the measurement $\mathbf{g}_i$ with a single observation $\mathbf{f}_i$ in the training set. To resolve this issue, we introduce the measurement process (dot-dashed line in Figure 3 (b)) to the signal retrieval process. The expected measurement, $\mathbf{g}^*$, is generated from $\boldsymbol{\mu}_\mathbf{f}^{(\theta)}(\mathbf{z}, \mathbf{g}_i)$ via the

forward process $\mathbf{g}^* = A\left(\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{z}, \mathbf{g}_i)\right)$. For all the $\mathbf{z}$ sampled from $p_\theta(\mathbf{z}|\mathbf{g}_i)$, we maximize the likelihood of generating the expected measurement $\mathbf{g}^*$ given point $\mathbf{g}_i$, as defined by the detection model. Applying Jensen's inequality, a lower bound of this likelihood $\mathcal{L}_r(\boldsymbol{\theta}; \mathbf{g}^*, \mathbf{g}_i)$ can be derived and used as the objective function of the retrieval process.

$$\log p(\mathbf{g}^*|\mathbf{g}_i) = \log \int p_\theta(\mathbf{g}^*|\mathbf{z}, \mathbf{g}_i) p_\theta(\mathbf{z}|\mathbf{g}_i)\, d\mathbf{z} \geq \int \log p_\theta(\mathbf{g}^*|\mathbf{z}, \mathbf{g}_i)\, p_\theta(\mathbf{z}|\mathbf{g}_i) d\mathbf{z} \tag{13}$$

$$= E_{p_\theta(\mathbf{z}|\mathbf{g}_i)}(\log p_\theta(\mathbf{g}^*|\mathbf{z}, \mathbf{g}_i))$$

$$\approx -\frac{1}{\alpha L} \sum_{l=1}^{L} \left(A\left(\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{z}_l, \mathbf{g}_i)\right) - \mathbf{g}_i\right)^2 \coloneqq \mathcal{L}_r(\boldsymbol{\theta}; \mathbf{g}^*, \mathbf{g}_i),$$

where we have assumed Gaussian noise model on the detector $\mathbf{g}^* \sim \mathcal{N}(\mathbf{g}_i, \alpha \mathbf{I})$. Notice that the Gaussian likelihood can be substituted with Poisson or binomial noise models in photon-limited detection [12,28]. The expectation $E_{p_\theta(\mathbf{z}|\mathbf{g}_i)}$ in Equation (13) is approximated by sampling L instances of from the conditional prior distribution $p_\theta(\mathbf{z}|\mathbf{g}_i)$ as $\{\mathbf{z}_l : l = 1, \ldots, L\}$. By adding in the measurement processes, we essentially construct a variational autoencoder for measurement $\mathbf{g}$, and the objective function promotes forward model consistency. The retrieval model can then be jointly trained alongside the inference model with a hybrid objective function [32]

$$\mathcal{L}_h(\phi, \theta, \mathbf{f}_i, \mathbf{g}_i) = \gamma \mathcal{L}(\phi, \theta; \mathbf{f}_i, \mathbf{g}_i) + (1 - \gamma)\mathcal{L}_r(\theta; \mathbf{g}^*, \mathbf{g}_i), \tag{14}$$

where the hyperparameter γ balances the weight between the two models.

CHAPTER 3: PHOTON-COUNTING COMPUTED X-RAY IMAGING

In this chapter we will apply the photon-counting statistics to X-ray imaging and discuss strategies to allocate a fixed photon budget for optimal region-of-interest reconstruction.

3.1 Verification and calibration of photon-counting noise model

We first verify the noise model of our photon-counting system and calibrate the incident photon flux λ with an X-ray projection measurement on a linear attenuation pattern, which was created by stacking multiple paper layers with identical thickness $h = 0.12$mm. The pattern was divided in to 3 X 3 regions, with region 1 being air and region 9 corresponding to 8 paper layers. We performed an X-ray-projection scan (Figure 4 (a1)) covering all 9 regions of the paper pattern, and waited for the arrival of the r=256th photon at each scanning step. The sampling interval of the detector Δt was10us. Figure 4 (a2) plots the average and standard deviation of the time intervals, $\mathbf{g}$, within each region in log scale. The signal-to-noise ratio, calculated as the ratio between the average and standard deviation fluctuates around 16.2, suggesting that the measurement uncertainty agrees with the shot-noise limit, the square root of photon counts. The linearity of the curve agrees with the exponential decay in the transmission as the thickness increases. From the slope in Figure 4 (a2), we estimated the transmittance, t=93% per paper layer. To directly observe the distribution of time intervals, we varied the number of photons, r, to collect at each point. Figure 4 (b1-b4) plot the histogram of elapsed intervals $\mathbf{g}$ within region 1 before the arrival of r=1, 2, 4 and 8 photons. We fit a negative binomial model with one unknown, *T*, on each

histogram. The red curves plot the negative binomial distributions with fitted parameter, T, as in Equation (2), which are 0.0127, 0.0128, 0.0129, and 0.0129 respectively in (b1)-(b4). The high consistency signifies the same photon flux exhibited on all histograms. The incident photon flux λ can then be calibrated from the T in region 1.

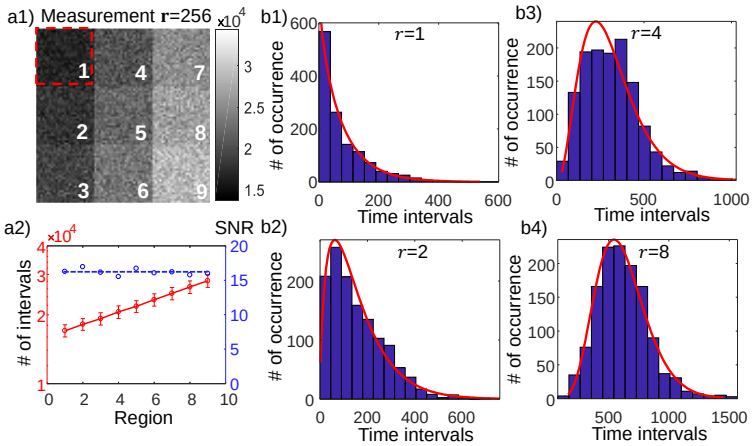

Figure 4: Experimental observation of the photon-counting model. (a1) Number of time intervals before the arrival of 256th photon. (a2) Time intervals in each region. Red dots and error bars represent the mean and standard deviation. Blue dots represent the signal-to-noise ratio (SNR) (b1-b4) Histograms of the number of time intervals elapsed before r=1, 2, 4, and 8 photons are detected in region 1.

3.2 Photon-counting CT imaging

With the calibrated photon flux, photon-counting CT scans were performed on an acrylic resolution target and a mouse brain sample. The results demonstrated comparable reconstruction quality and reduced dose as conventional, time-integrating CT scans.

Figure 5 shows the experimental results on a resolution target. A reference image was collected by measuring the photon counts within 1s (Figure 5 (a1), $\mathbf{g}$=10^5). The average

number of photons per beam was 569. The reconstructed image (Figure 5 (b1)) was used as ground truth for evaluating low-photon-count images. Figure 5 (a2) and (a3) present two different low-photon-count measurements, one with constant acquisition intervals (a2, $\mathbf{g}$=6250) and counts the number of detected photons, the other waits for constant photon counts (a3, $\mathbf{r}$=16) and records the number of elapsed time intervals, for all the pencil beam measurements. The average detected counts per pencil beam in (a2) was 17.8, comparable to the counts in (a3). The reconstructed attenuation maps are shown in Figure 5 (b2, b3), respectively. The intensity profile of 0.7mm group is plotted in Figure 5 (c), which shows a visibility of 0.82 on reference image, and 0.68, 0.60 for the reconstructed images in (b2) and (b3), respectively. The spatial resolution is limited by the 0.6mm spot size on the sample plane due to the beam divergence. With approximately the same average photon count, Figure 5 (b2) and (b3) have normalized mean square difference of 6.2% and 4.9% with respect to the reference image. We speculate that the smaller NMSE in Figure 5 (b3) is primarily attributed to the uniform SNR on the entire sinogram (Figure 5 (a3)), whereas in Figure 5 (a2), the central region of the sinogram suffers from low SNR.

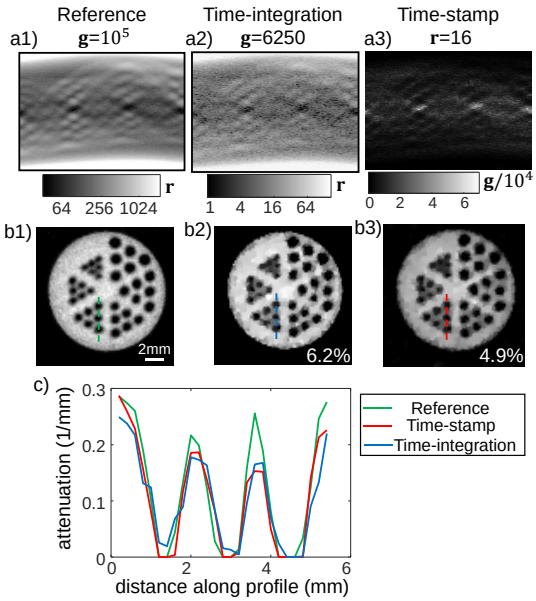

Figure 5 Measurement (a) and reconstruction (b) of 1) reference image, 2) time-integration PC-CT scheme (17.8 photons/beam on average) and 3) time-stamp PC-CT (16 photons/beam) of a resolution target. (c) attenuation profile along the dashed line (0.7mm line-width group) in (b1-b3).

Photon counting CT has the potential in radiation dose reduction, which is especially attractive for biomedical imaging applications. We compared the image of a mouse brain layer obtained from CT scans using photon-counting detector (PC-CT for short) and a flat panel detector (FPD-CT, for short). Figure 6 shows the reconstruction from FPD-CT (a, 12kV 40mA source power, 0.5s integrating time, 116.2 detector readout per beam on average) and time-stamp PC-CT (b, 16 photons per beam, 12kV 4mA source power). A comparison on the absorbed radiation doses between Figure 6 (a) and (b) was performed through Monte Carlo simulation. The irradiance of the source was calculated using *XSPECT* under experimental power settings. The radiation dose of time-stamp PC-CT was calculated via an equivalent tube current modulation to simulate different integration time for each pencil beam with *ImpactMC* [24]. Figure 6 (c) plots the

percentage of PC-CT radiation dose with respect to that in FPD-CT. The proposed PC-CT reduces the dose to ~0.6% of FPD-CT, because the photon-counting detector eliminates the dark noise commonly found on panel detectors, and thus a much lower source flux could be used for image acquisition. Figure 6 (c) shows that the dose reduction on the surface is more prominent than the interior region. This is because the transmission signal from the interior region is weaker, and the photon-counting scheme would wait longer when scanning in the interior region until the pre-defined photon counts are received.

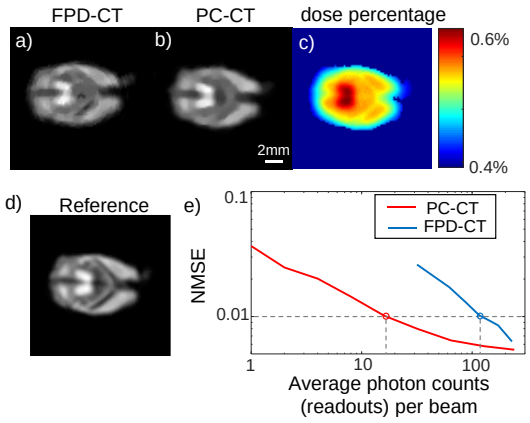

Figure 6: Mouse brain sample imaged with (a) panel detector (FPD-CT, average 116.2 detector readout/beam) and (b) time-stamp PC-CT (16 photons/ beam). (c) Radiation dose ratio between time-stamp PC-CT and FPD-CT. (d) Reference image with 1s integration time per pencil beam (e) NMSE versus average photon counts per beam for time-stamp PC-CT and panel detector.

To further evaluate the performance between PC-CT and FPD-CT, we acquired a complete photon time stamp spanning 1s integration time (1283 photon counts per beam on average) for the mouse brain sample, and reconstructed a reference image (Figure 6 (d)) from all the detected photons. Figure 6 (e) plots the normalized mean square difference between the reconstruction and the reference in log scale. The blue

and red circles on the plot correspond to FPD-CT and PC-CT in Figure 6 (a) and (b), respectively. For low photon counts, PC-CT consistently performs better than conventional CT. As the photon count increases, flat-panel detector eventually will have a comparable reconstruction error as that of PC-CT. This is because, in high-photon flux regime, the noise model of using the panel detector and photon counting module can both be approximated by a Gaussian distribution.

3.3 Photon-allocation strategy for region-of-interest CT imaging

In medical or industrial CT scan, the ultimate imaging task is often diagnosis or detection of a specific feature within a region-of-interest (ROI), rather than the whole volume reconstruction [25,26]. Task-specific image acquisition design can potentially shorten imaging time and reduce radiation damage to the region irrelevant to the diagnostic goal. Conventionally the image acquisition design for ROI reconstruction follows a scheme called “interior tomography” [2], which distributes the entire radiation dose budget exclusively to the ROI, resulting in a series of truncated projections. A unique and stable ROI reconstruction becomes a mathematical challenge. Another acquisition strategy is to use low-resolution projection from the exterior region to stabilize the ROI reconstruction [25,27]. This approach can be considered as a trade-off between reconstruction stability of whole CT scan and dose reduction benefit of truncated scan. Yet this trade-off has not been quantitatively studied, mainly because the illumination or integration time of each pencil beam is not easily adjustable in a conventional setup [28]. The photon-counting framework provides an opportunity to

study and design the strategy of allocating a fixed photon budget for an imaging-specific task, mainly the ROI reconstruction. Since the measurement is a random process under the presence of noise, the estimated $\hat{\mathbf{f}}$ is also a random variable, whose bias and variance dictates the error of the reconstructed image/signal of interest. This section derives the error of the estimator under different photon-allocation designs and discusses the strategy for optimal region-of-interest (ROI) reconstruction performance.

To facilitate derivations on the distribution of the estimator $\hat{\mathbf{f}}$, we introduce auxiliary variables

$$\mathbf{t} = \left\{ t_j(\mathbf{f}) = \log \frac{\lambda}{T_j(\mathbf{f})} = \sum_{i=1}^{n} A_{ji} f_i \, , j = 1,2, \dots, m \right\}, \tag{15}$$

which can be interpreted as the CT line-integrals. By setting the derivative of the negative binomial measurement (Equation (2)) with respect to T_j to 0 and applying the invariance principle, the maximum likelihood (ML) estimation for t_j is achieved at

$$\hat{t}_j = \log \frac{\lambda r_j}{g_j}. \tag{16}$$

A change of variable in the distribution (Equation (2)) from $\mathbf{g}$ to $\hat{\mathbf{t}}$, and a Taylor-expansion on the resulting $\hat{\mathbf{t}}$ distribution to the second order around $\mathbf{t}$ gives a normal distribution that each element j in $\hat{\mathbf{t}}$ approximately follows $\hat{t}_j \sim \mathcal{N}(\mu_{\hat{t}_j}, \sigma^2_{\hat{t}_j})$, where $\mu_{\hat{t}_j}$ equals the ground truth of CT line integral t_j, and the variance $\sigma^2_{\hat{t}_j} = 1/r_j$ is inversely proportional to the photon count received at that pencil beam [29]. Next, we apply Taylor-expansion to the negative binomial likelihood in Equation (7) around the

estimated line-integral $\hat{\mathbf{t}}$ (Equation (16)), and reduce the MAP estimator in Equation (5) to a least-square problem weighted on photon count $\mathbf{r}$ [29]

$$\begin{aligned}\hat{\mathbf{f}}(\hat{\mathbf{t}};\mathbf{r}) &= \underset{\mathbf{f}'}{\operatorname{argmin}}\{L(\mathbf{f}'|\mathbf{g};\mathbf{r})\} \\ &\approx \underset{\mathbf{f}'}{\operatorname{argmin}}\{\frac{1}{\mathbf{2}}\left\|diag(\sqrt{\mathbf{r}})(\mathbf{A}\mathbf{f}'-\hat{\mathbf{t}})\right\|_2^2+\tau\|\mathbf{D}\mathbf{f}'\|_2^2\},\end{aligned} \tag{17}$$

where the zero-order terms in y_j are independent of $\mathbf{f}$, and can thus be neglected in the optimization; $diag(\sqrt{\mathbf{r}})$ denotes the diagonal matrix constructed from the vector $\sqrt{\mathbf{r}}$. The resulting objective function in Equation (17) $\varepsilon(\mathbf{f})=\frac{1}{\mathbf{2}}\left\|diag(\sqrt{\mathbf{r}})(\mathbf{A}\mathbf{f}-\hat{\mathbf{t}})\right\|_2^2+\tau\|\mathbf{D}\mathbf{f}\|_2^2$ has a gradient

$$\nabla\varepsilon(\mathbf{f})=\mathbf{A}^T diag(\mathbf{r})(\mathbf{A}\mathbf{f}-\hat{\mathbf{t}})+2\tau\mathbf{D}^{\boldsymbol{T}}\mathbf{D}\mathbf{f} \tag{18}$$

and Hessian matrix

$$\mathbf{H}(\mathbf{r})=\mathbf{A}^T diag(\mathbf{r})\mathbf{A}+2\tau\mathbf{D}^{\boldsymbol{T}}\mathbf{D}. \tag{19}$$

For a properly chosen regularization matrix $\mathbf{D}$, $\mathbf{H}$ is positive-definite and can thus be conceived as the inverse covariance of the posterior distribution $p(\mathbf{f}|\,\hat{\mathbf{t}})$. Equation (17) has an explicit solution similar to a Tikhonov regularization [30] weighted on $\mathbf{r}$

$$\hat{\mathbf{f}}(\hat{\mathbf{t}};\mathbf{r})=\mathbf{H}^{-1}(\mathbf{r})\mathbf{A}^T diag(\mathbf{r})\hat{\mathbf{t}}, \tag{20}$$

where $\mathbf{H}^{-1}(\mathbf{r})$ denotes the inverse of Hessian matrix $\mathbf{H}(\mathbf{r})$. Since the estimator $\hat{\mathbf{f}}$ is a linear superposition of Gaussian variables $\hat{\mathbf{t}}$, the distribution of $\hat{\mathbf{f}}$ thus follows a multivariate Gaussian distribution with mean

$$\boldsymbol{\mu}_{\hat{\mathbf{f}}}(\mathbf{r};\mathbf{f}) = \mathbf{H}^{-1}(\mathbf{r})\mathbf{A}^{T} diag(\mathbf{r})\mathbf{A}\mathbf{f} \tag{21}$$

and covariance

$$\boldsymbol{\Sigma}_{\hat{\mathbf{f}}}(\mathbf{r}) = \mathbf{H}^{-1}(\mathbf{r})\mathbf{A}^{\boldsymbol{T}} diag(\mathbf{r})\mathbf{A}\mathbf{H}^{-1}(\mathbf{r}). \tag{22}$$

The MSE of the estimator consists of the bias square and the variance of the pixels within ROI

$$MSE_{\hat{\mathbf{f}}}(\mathbf{r};\mathbf{f}) = \|(\boldsymbol{\mu}_{\hat{\mathbf{f}}}(\mathbf{r};\mathbf{f}) - \mathbf{f})\mathbf{w}\|^2 + tr[diag(\mathbf{w})\boldsymbol{\Sigma}_{\hat{\mathbf{f}}}(\mathbf{r})]. \tag{23}$$

Equation (23) provides an analytical expression of the MSE of the estimator with respect to $\mathbf{r}$, the number of photons to wait at each pencil beam. For a given total photon budget I_0, the optimal strategy minimizes the ROI reconstruction MSE

$$MSE = \mathbf{w} \odot MSE_{\hat{\mathbf{f}}}(\mathbf{r};\mathbf{f}), \tag{24}$$

subject to

$$\sum_j r_j = I_0, \tag{25}$$

where the weight $\mathbf{w}$ is 1 inside ROI, and 0 outside ROI; $\odot$ represents elementwise product.

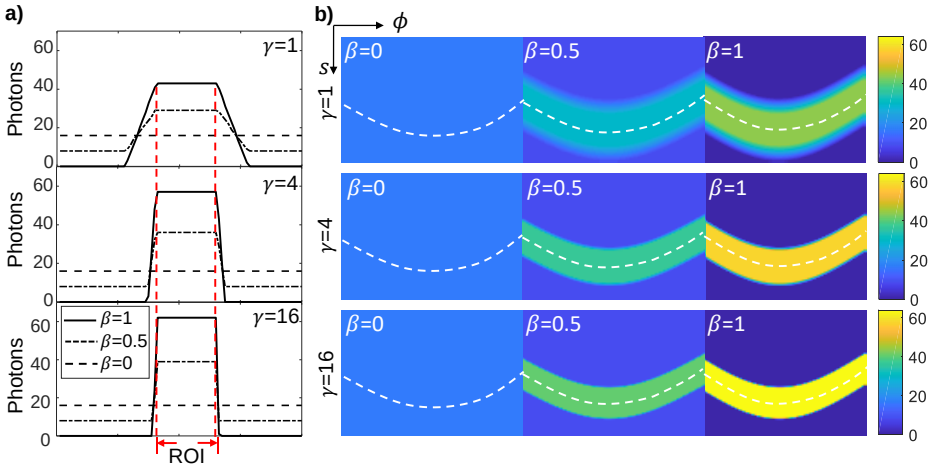

Figure 7: Photon allocation strategies parameterized by the parameter β and γ. (a) Photon allocation profiles at γ=1, 4 and 16 under 0° projection. (b) two-dimensional photon allocation maps for β=0, 0.5 and 1, γ=1, 4 and 16.

Because it is impossible to enumerate all possible photon allocation strategies $\mathbf{r}$, we confine our choice of the photon allocation strategies to a trapezoid function in the translation and rotation dimensions, which can be expressed by

$$r_{j_s,j_\phi} = \frac{I_0(1-\beta)}{m_\phi} + \begin{cases} \frac{1}{m_\phi}\frac{I_0\beta\delta x}{2\sigma+\Delta}, & |j_s\delta x - s_c| \le \sigma \\ \frac{1}{m_\phi}\frac{I_0\beta\delta x}{2\sigma+\Delta}\frac{\sigma+\Delta-|j_s\delta x - s_c|}{\Delta}, & \sigma < |j_s\delta x - s_c| < (\sigma+\Delta) \\ 0, & |j_s\delta x - s_c| \ge (\sigma+\Delta) \end{cases} \quad (26)$$

$$j_s = 1,2,\dots,m_s$$

$$j_\phi = 1,2,\dots,m_\phi$$

where δx is the translation step size; σ is the radius of ROI; β controls the interior/exterior ratio; Δ denotes the width of the transition region where the photon

number drops from maximum to the minimum; when Δ=0, the photon allocation profile becomes a rectangular shape. In our simulation, we define $\gamma = \sigma/\Delta$ as the shape parameter that describes the slope of the photon allocation profile across the ROI boundary. s_c is the center coordinate of ROI at different projections, and is calculated via

$$s_c = s_{offset}\sin(\phi + \phi_{offset}), \tag{27}$$

where s_{offset} is the offset between the ROI center and the rotation center; ϕ_{offset} is the azimuthal coordinate of the ROI center with respect to the rotation center. Figure 7 plots the photon allocation profile along the translation direction, s, under 0° projection, as well as the 2D photon map when different projections angles are considered.

Comprehensive simulation results in Ref. [29] have shown that the reconstruction MSE in ROI region mainly depends on the β parameter, the portion of photon budgets allocated to the ROI, when the shape parameter γ is larger than 1. The optimal ratio is primarily determined by the size of ROI region. Strategies with $\gamma = 1$ is not preferred as it would allocate too many photons to the vicinity of ROI, reducing the overall SNR in ROI. The “interior tomography” scheme can be incorporated into our framework as a special case $\beta = 1$, $\gamma \to \infty$. We found γ=16 is large enough to avoid allocating any photons outside ROI.

Since it is possible to analytically predict the ROI reconstruction error, we tested our strategy on an experimental phantom in Figure 8. Figure 8(a) shows a full-scan CT image of an acrylic resolution target acquired with 1s integration time and an average

photon count of 589 per pencil beam, which was used as the reference. The ROI reconstruction was performed at various photon allocation strategies with 16 photons per beam on average. For each strategy, we calculated the bias and variance with 6 regularization parameters τ ranging from $10^1 \sim 10^{3.5}$, and selected the one giving minimal predicted MSE for use in the SPIRAL reconstruction. Figure 8 (b) plots the analytically predicted reconstruction MSE (compared to the reference in (a)) inside the ROI as a function of interior/exterior ratio β and the shape parameter γ. From our prediction, the smallest ROI reconstruction error was attained at β=0.7, γ=4 with the regularization parameter τ=10^2. Figure 8 (d, e) shows the measured time intervals $\mathbf{g}$ (d1-d3) and reconstructions (e1-e3) from 3 photon allocation maps $\mathbf{r}$ (c1-c3), corresponding to uniform (c1), optimized (c2) and interior (c3) strategies, respectively. The optimized strategy had a reconstruction NMSE of 2.8% at ~30 times reduced photon budget compared to the reference image.

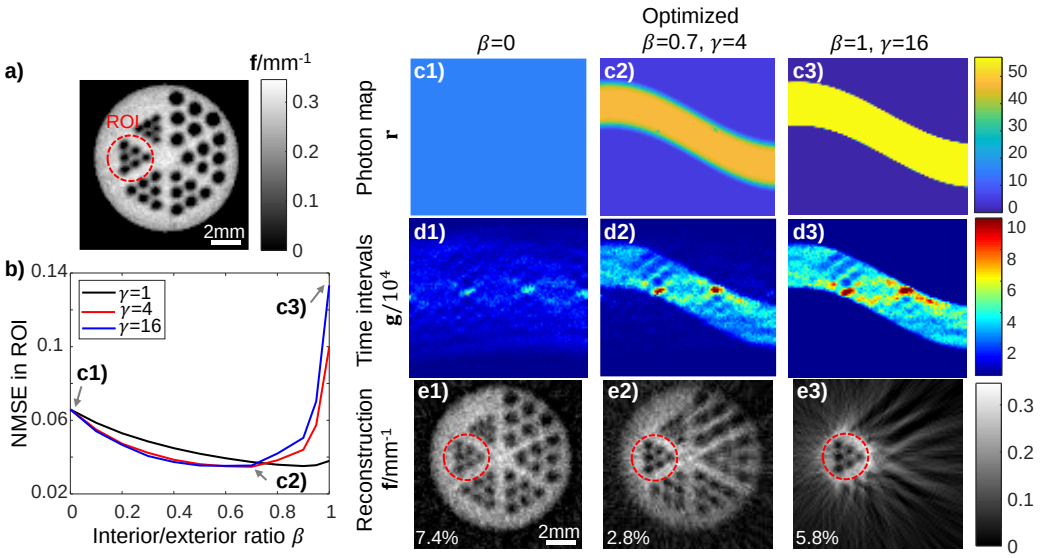

Figure 8: Comparison between different photon allocation strategies for the resolution target. (a) Full-scan CT image of the resolution target. The ROI covers the 0.6mm line-width group. (b) Predicted reconstruction MSE in ROI with respect to different photon allocation strategies, expressed in terms of the interior/exterior ratio β and the shape parameter γ. (c–e) Examples of the photon allocation strategies, experimental measurements and the corresponding ROI reconstructions from (1) uniform photon counts (2) optimized photon allocation map (3) interior measurement with β=1, γ=16. The average photon count was 16 per beam. All scale bars represent 2mm. The ROI is marked by the red, dashed circle in (a) and (e). The numbers in the left bottom of each reconstruction indicate the MSE within ROI.

CHAPTER 4: STATISTICAL ATTOSECOND PULSE RETRIEVAL

The key to statistical image/signal reconstruction is finding the posterior distribution, which describes the distribution of all plausible images/signals given a specific measurement. In the previous chapters we have demonstrated analytical derivation of the posterior distribution and construction of MAP estimator, for a linear measurement process. In this chapter we will apply learning-based approach, namely "conditional variational generative network" (CVGN), to capture this posterior distribution for a nonlinear measurement process, attosecond pulse retrieval from noisy streaking traces, which does not have a closed form of posterior distribution. Conventional attosecond pulse retrieval has been solved with FROG-CRAB [3] and its derived algorithms. However, these algorithms have made a series of assumptions described below, which prevented them from reconstructing pulses shorter than atomic unit of time [31]. We will demonstrate the capability of CVGN that can incorporate a complete attosecond streaking process and model the reconstruction uncertainties, both of which are desirable capabilities not yet possessed by conventional FROG-CRAB-derived algorithms.

4.1 Attosecond streaking process

Here we consider the complete physical process of attosecond streaking. The streaking trace is a series of photoelectron spectra $y(|\vec{k}|, \tau)$ generated from the interaction between an attosecond XUV pulse $E_{XUV}(t)$ and a femtosecond infrared (IR) pulse $E_{IR}(t)$ in a gaseous medium with various time delays τ. Both XUV and IR pulses are polarized

along the axis of time-of-flight (TOF) spectrometer. Once the gas atoms with ionization potential I_p absorb the energy from an XUV pulse, photoelectrons with momentum $\vec{k}$ (and hence kinetic energy $W = \left|\vec{k}\right|^2/2$) are produced via dipole transition $d(\vec{k})$. The dressing IR field introduces a momentum shift to the photoelectrons, which is manifested in the form of a phase modulation, ϕ_G, on the photoelectron wave packet. The spectrogram $S(\vec{k}, \tau)$ is described by Equation (28) in atomic units

$$S\left(\vec{k}, \tau, \theta\right) = \left| \int_{-\infty}^{\infty} E_{XUV}(t - \tau) d_{\vec{k}} \exp\left(\mathrm{i}\phi_G\left(\vec{k}, t\right)\right) \exp\left(\mathrm{i}\left(W + I_p\right)t\right) dt \right|^2, \tag{28}$$

$$\phi_G\left(\vec{k}, t, \theta\right) = -\int_{t}^{\infty} (\left|\vec{k}\right| A(t') \cos\theta + |A(t')|^2/2) dt', \tag{29}$$

where $A(t) = -\int_{-\infty}^{t} E_{IR}(t')dt'$ is the vector potential of the IR field along the TOF spectrometer; θ is the observation angle between the photoelectron momentum, $\vec{k}$, and the TOF axis. If we denote the angular distribution of the photoelectron against the TOF axis as $f(\theta)$, the ideal streaking trace, I, expressed in terms of photoelectron energy W and delay τ between XUV and dressing IR pulse, is the integral over all the directions of $\vec{k}$ that fall within the maximum collection angle θ_{max} of the TOF spectrometer

$$I(W, \tau) = \int_{\theta=0}^{\theta_{max}} S\left(\vec{k}, \tau, \theta\right) f(\theta) \sin(\theta)\, d\theta, \tag{30}$$

where

$$f(\theta) = 1 + \frac{\beta}{2}(3\cos^2\theta - 1), \tag{31}$$

where β is a parameter that accounts for the asymmetry distribution of the photoelectrons ionized from different gas atoms [32]. Experimentally obtained traces typically contain shot noise due to the low photoelectron flux. If the ideal trace (Equation (30)) is normalized by its maximum intensity, and the experimental average peak count of the trace is λ, for each energy channel W and delay τ, the measured count y on the trace follows Poisson statistics

$$p(y|I) = \frac{(\lambda I)^y \exp(-\lambda I)}{y!}, \tag{32}$$

where the parameter λ is determined by the photoelectron flux and the integration time. In most FROG-derived pulse retrieval methods, the dipole transition element, d, is assumed to be constant. This assumption does not hold for low energy photoelectrons, which are subject to additional energy-dependent group delays from Coulomb-laser coupling [33,34]. This delay can be expressed as a phase term

$$d(W) = \exp(\mathrm{i}\eta(W))\exp(\mathrm{i}\delta(W)), \tag{33}$$

where the energy W is expressed in atomic unit; $\eta(W) = \arg\{\Gamma(2 - \mathrm{i}/\sqrt{2(W - I_p)})\}$ is the phase delay from the Coulomb potential [34], $\Gamma[\cdot]$ representing the complex gamma function; $\delta(W) = \int_0^{W-I_p}(2W')^{3/2}(2 - \ln(W'T_{IR}))dW'$ is the phase delay from the interaction with IR dressing field [33], where $T_{IR} = 2\pi c/\lambda_c$ is the oscillating period of the IR field in atomic unit.

4.2 CVGN training and pulse retrieval from attosecond streaking trace

To construct the dataset for training the CVGN, XUV and IR pulses were represented by a set of parameters, $\mathbf{x}$. Because the spectral density of the XUV pulse, $S_{XUV}(\epsilon)$ was measured experimentally, the XUV field can be uniquely determined by its spectral phase $\phi_{XUV}(\epsilon)$, which was expressed as a 5th order Taylor series in our model,

$$\tilde{E}_{XUV}(\epsilon) = \sqrt{S_{XUV}(\epsilon)}\exp\mathbb{i}\sum_{i=1}^{5} k_i\epsilon^i. \tag{34}$$

The XUV pulse in spectral domain were Fourier transformed into time-domain for streak calculation. In addition, 4 elements representing the carrier envelop phase, central wavelength, pulse duration and peak intensity [35] were used to characterize the IR dressing pulse, totaling the number of XUV and IR pulse parameters to 9.

The training data were generated by adding noise to the ideal, noise-free traces from the physics model. We first created 10000 ideal traces with random pulse parameters, $\mathbf{x}$, and normalized their intensities to the range between 0 and 1. Poisson noise was added to each ideal trace $\mathbf{y}_0$ to simulate noisy traces in experiments

$$\mathbf{y} \sim Poisson(\lambda\mathbf{y}_0), \tag{35}$$

where the parameter λ is the average peak count of the Poisson distribution. The training data contained a mixture of 5 different Poisson noise levels, λ=10, 32.5, 55, 77.5 and 100. An additional 1000 ideal traces were used to test the trained model, which consisted of 10 different Poisson noise levels, λ, ranging from 5 to 100.

The structure of CVGN used for attosecond streaking trace retrieval was constructed according to that in Ref. [22]. It is worth noting that the output (label) of this CVGN $\mathbf{x}$ is a vector with the 9 pulse parameters, instead of the real and imaginary part of the XUV spectrum in Ref. [22]. After the training, 25 instances of pulse parameters $\hat{\mathbf{x}}$ were retrieved from the distribution $p_\gamma(\mathbf{x}|\mathbf{z},\mathbf{y})$ by sampling $\{\mathbf{z}_l: l = 1, \dots, 25\}$ from the conditional prior distribution $p_\gamma(\mathbf{z}|\mathbf{y})$. The time- and frequency-domain XUV pulses were then reconstructed from these pulse parameters.

The retrieved XUV pulses from both test dataset and experimental streaking traces using a trained CVGN are presented in Figure 9 to show the accuracy at various noise levels. Figure 9 (a1-a3) shows three simulated streaking traces from the same noise-free test trace with λ=5, 21 and 100. For each trace, 25 instances of the pulse parameter set, $\hat{\mathbf{x}}$, were retrieved from the posterior distribution $p_\gamma(\mathbf{x}|\mathbf{y},\mathbf{z})$. The frequency (Figure 9 (b1-b3)) and time-domain (Figure 9 (c1-c3)) XUV pulses were then reconstructed from these retrieved pulse parameters, and streaking traces (Figure 9 (d1-d3)) were generated from the retrieved XUV pulses using the physical model. Error bars on Figure 9 (b, c) represent the variance of the 25 instances of retrieved XUV pulse.

The reconstructed streaking traces and the full-width-at-half-maximum (FWHM) of the time-domain XUV pulse were compared with the ground truth. The mean squared error (MSE) of the 25 reconstruction instances was used as the error metric to evaluate the performance under various noise levels. The reconstructed and ideal streaking traces

were all normalized to facilitate the comparison among various λ. For the errors on pulse duration, the MSE was normalized by the FWHM of the XUV ground truth. Figure 10 (a, b) plot the MSE on the reconstructed streaking traces and pulse FWHM with respect to λ, averaged over the whole test dataset.

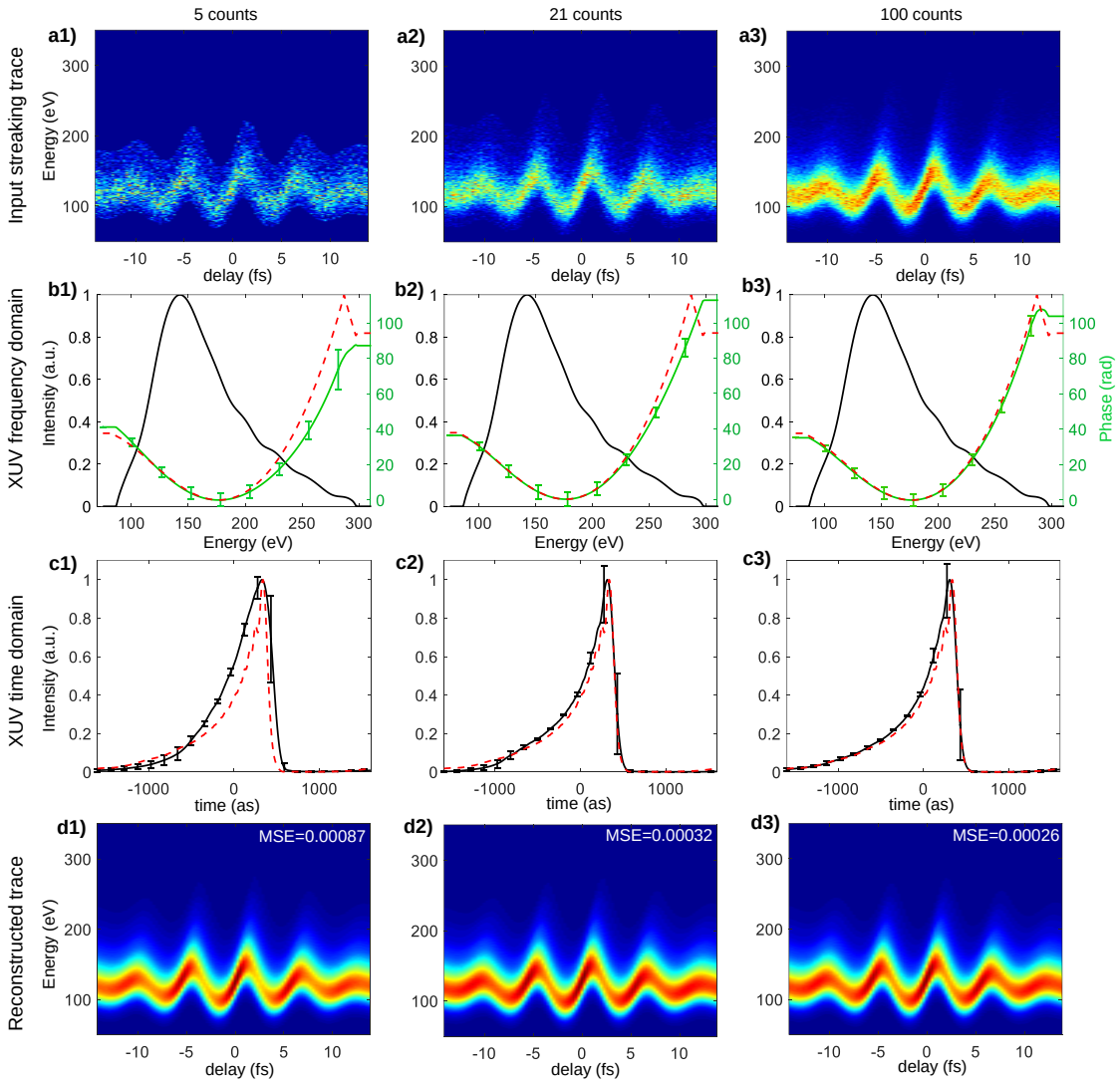

Figure 9: Pulse retrieval from a simulated streaking trace. (a) Input streaking traces. (b, c) Retrieved XUV pulse in (b) frequency domain and (c) time domain. The dashed red curves on (b) and (c) indicate the ground truth of time-domain XUV pulse and spectral phase. Error bars on (b), (c) represent the variance of the retrieved pulse instances. (d) Reconstructed streaking trace from the retrieved pulses. The MSE indicates the error between the reconstructed streaking trace and the ground truth.

Results from Figure 10 show that as the average peak count, λ, exceeds 32.5, the MSE of the reconstructed streaking trace decreased and remained below 3.5×10^{-4}. The MSE of pulse duration also decreased from ~80% (λ=5) down to 6% ($\lambda = 32.5$). The results suggest an average peak Poisson SNR of at least 6 to achieve satisfactory pulse retrieval. It is worth noting that for low photon count (5 counts), the retrieved instances showed increased bias in pulse profile, which is an indication of strong regularization [29]. This is the effect of using the training data with mixed noise level, which can be reduced by training with traces with same noise level or implementing additional mechanism to adjust the posterior distribution based on noise level.

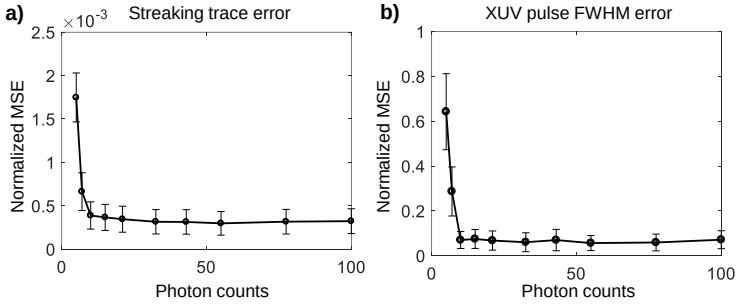

Figure 10: Normalized MSE of (a) reconstructed streaking trace (b) FWHM of retrieved time-domain XUV pulse at various Poisson noise levels. The error bars represent the MSE fluctuation within the whole test dataset.

4.3. Resolving carrier phase ambiguity in attosecond streaking process

Using the knowledge of the streaking process, CVGN can generate instances that satisfy the measurement process. This could be a useful feature in exploring the ambiguities of the measurement process, in which multiple signals/images that look drastically different correspond to the same observation due to the ill-conditioned

measurement process. From Equation (28), it is clear that pulses differ only in the carrier envelop phase, k_0, will produce the same measurement. This carrier phase ambiguity cannot be easily resolved in conventional neural networks with a deterministic structure and will adversely affect the performance of those deterministic neural networks.

For the retrieval of ambiguity instances, we additionally generated 10 ambiguity instances of each ideal streaking trace by changing the k_0 term within 0 and 2π, based on the 10000 traces in the training dataset. As a comparison, we also constructed a deterministic neural network by combining the encoding $\boldsymbol{\mu}_{\mathbf{z}}^{(\theta)}(\mathbf{g})$ and decoding $\boldsymbol{\mu}_{\mathbf{f}}^{(\theta)}(\mathbf{g},\mathbf{z})$ structures in CVGN, and removing the sampling process $\mathbf{z}\sim q_{\phi}(\mathbf{z}|\mathbf{f}_i,\mathbf{g}_i)$. Both deterministic neural network and CVGN were trained on this dataset that contained ambiguity.

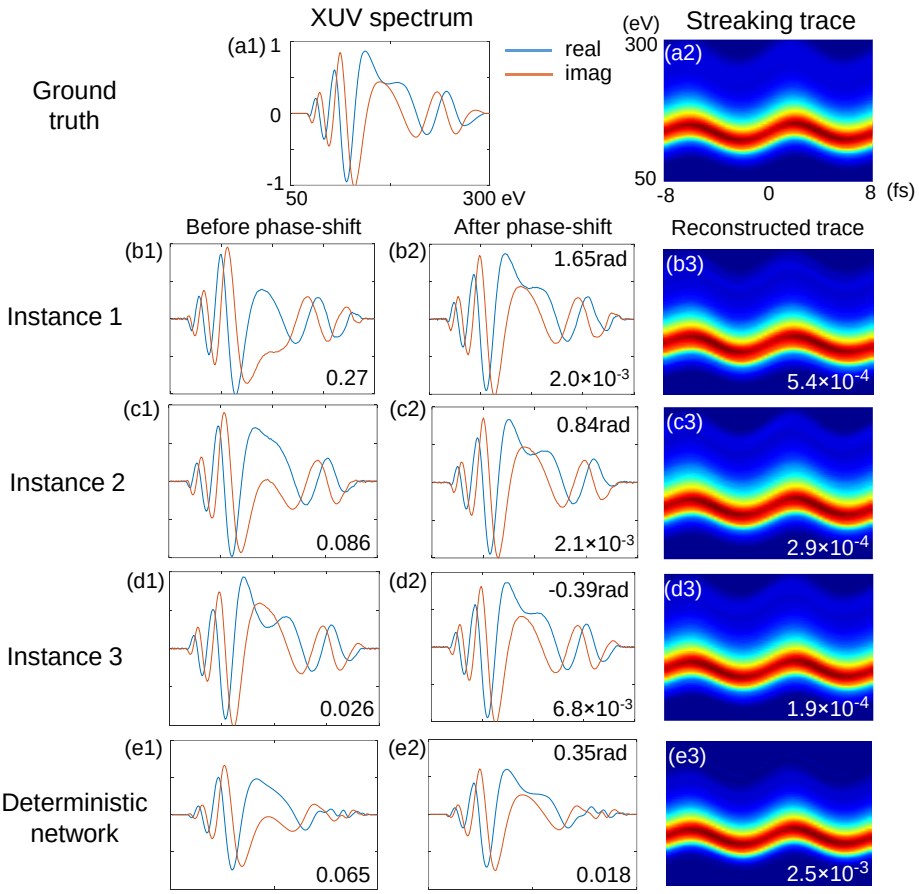

Figure 11: Reconstructions from the ultrafast pulse retrieval experiment: (a) Ground truth of the real and imaginary part of the XUV spectrum and its simulated streak trace. The IR spectrum is not shown in the figure. (b-d) Three instances of CVGN-retrieved XUV spectrum (b1-d1), their phase-shifted variant (b2-d2), and the streaking trace (b3-d3) calculated from each instance. (e) Retrieved XUV spectrum, its phase-shifted variant and streak trace from the deterministic network.

Figure 11 displays the real and imaginary part of the CVGN-retrieved XUV spectrums to highlight the differences among ambiguity instances. An ideal streaking trace (Figure 11 (a2)) generated from a test XUV pulse (Figure 11 (a1)) was fed into the trained CVGN. Three instances of the CVGN-retrieved XUV spectrums are shown in Figure 11 (b1-d1), with MSE of 0.27, 0.086 and 0.026, respectively, compared with the ground truth in Figure 11 (a). Figure 11 (b3-d3) show the traces reconstructed from retrieved pulses (b1-d1), and the fidelity with respect to the ground truth in Figure 11 (a2).

The high measurement fidelities suggest that instances in Figure 11 (b1-d1) belong to the phase-shift ambiguities of the same streaking trace. To verify this, we shifted the carrier envelop phase k_0 by the average phase difference within 100~300eV. The resulting pulse spectra (Figure 11 (b2-d2)) match the retrieved XUV spectrum and ground truth with good consistency. The amount of phase-shift was 1.65, 0.84 and -0.39 radians, respectively for Figure 11 (b2-d2), with MSE of 2.0×10^{-3}, 2.1×10^{-3} and 6.8×10^{-3}after the phase shift. In contrast, the deterministic network generates identical reconstructions similar to the average of all ambiguity instances. The XUV spectrum in Figure 11 (e1) cannot be phase-shifted to match the ground truth and exhibits poor fidelity (Figure 11 (e3)) compared with the true streaking trace. Table 1 summarizes the average MSE of the retrievals from the test dataset. CVGN demonstrates good fidelity as it resolves the individual ambiguity instances, a capability unmatched by deterministic network. To reach similar retrieval fidelity, a deterministic network requires manually removing the ambiguity instances from the training data.

Table 1: MSE of retrieved pulses and reconstructed traces using deterministic network and CVGN

MSE of	Deterministic network	CVGN
Pulse	7.8×10^{-3}	4.1×10^{-3}
Trace	4.9×10^{-3}	7.1×10^{-4}

CHAPTER 5: CONCLUSION

In this dissertation, we have established a statistical estimation framework for computational image reconstruction from photon-counting measurements. The statistical image reconstruction process models the posterior distribution of the plausible images/signals given an observed measurement, and outputs the solution, or multiple solutions in case of measurement ambiguity, with the highest likelihood. The presented framework incorporates the image-formation process and noise statistics of the imaging system to produce reconstructed images/signals that are consistent with the measurements. This versatile approach can be applied to estimate images/signals of interest in a wide variety of problems including photon-counting CT, region-of-interest tomography, and attosecond pulse retrieval. The statistical reconstruction framework has demonstrated capabilities unmatched by conventional reconstruction algorithms. In X-ray CT, high fidelity image reconstructions can be performed with as low as 16 photons per pixel, significantly reducing the radiation exposure to the sample compared with conventional CT system. We have also derived the optimal photon allocation strategy based on the MSE of the estimator. In attosecond pulse retrieval, the statistical retrieval framework demonstrates an SNR of 6 or better for satisfactory pulse retrieval results, which provides a guideline to future attosecond streaking experiments. Moreover, estimating all the possible pulses given an observed trace resolves the carrier phase ambiguity intrinsic to the streaking process. We envision that further development on the learning-based framework can enable fast, high-fidelity

image/signal reconstruction with the full inclusion of the physical model in the conditional variational generative network.

LIST OF REFERENCES

1. A. Kirmani, D. Venkatraman, D. Shin, A. Colaço, F. N. C. Wong, J. H. Shapiro, and V. K. Goyal, "First-Photon Imaging," Science (80-.). **343**, 58–61 (2014).

2. Z. Zhu, A. Katsevich, A. J. Kapadia, J. A. Greenberg, and S. Pang, "X-ray diffraction tomography with limited projection information," Sci. Rep. **8**, 522 (2018).

3. Y. Mairesse and F. Quéré, "Frequency-resolved optical gating for complete reconstruction of attosecond bursts," Phys. Rev. A - At. Mol. Opt. Phys. **71**, 1–4 (2005).

4. H. Wang, M. Chini, S. D. Khan, S. Chen, S. Gilbertson, X. Feng, H. Mashiko, and Z. Chang, "Practical issues of retrieving isolated attosecond pulses," J. Phys. B At. Mol. Opt. Phys. **42**, (2009).

5. M. Weik, R. B. G. Ravelli, G. Kryger, S. McSweeney, M. L. Raves, M. Harel, P. Gros, I. Silman, J. Kroon, and J. L. Sussman, "Specific chemical and structural damage to proteins produced by synchrotron radiation," Proc. Natl. Acad. Sci. **97**, 623–628 (2000).

6. R. B. Ravelli and S. M. . McSweeney, "The 'fingerprint' that X-rays can leave on structures," Structure **8**, 315–328 (2000).

7. D. Bernard and R. C. Blish, "Considerations for Minimizing Radiation Doses to Components during X-ray Inspection," in *2005 7th Electronic Packaging Technology Conference* (IEEE, 2005), Vol. 2, pp. 697–704.

8. A. Ditali, M. Ma, and M. Johnston, "X-Ray Inspection-Induced Latent Damage in DRAM," in *2006 IEEE International Reliability Physics Symposium Proceedings* (IEEE, 2006), pp. 266–269.

9. Z. Zhu, R. A. Ellis, and S. Pang, "Coded cone-beam x-ray diffraction tomography with a low-brilliance tabletop source," Optica **5**, 733–738 (2018).

10. M. Hassan, J. A. Greenberg, I. Odinaka, and D. J. Brady, "Snapshot fan beam coded aperture coherent scatter tomography," Opt. Express **24**, 18277 (2016).

11. Z. Zhu, A. Katsevich, and S. Pang, "Interior x-ray diffraction tomography with low-resolution exterior information Interior x-ray diffraction tomography with low-resolution exterior information," (2019).

12. K. Taguchi and J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," Med. Phys. **40**, 100901 (2013).

13. A. Markman and B. Javidi, "Integrated circuit authentication using photon-limited x-ray microscopy," Opt. Lett. **41**, 3297 (2016).

14. A. McCarthy, N. J. Krichel, N. R. Gemmell, X. Ren, M. G. Tanner, S. N. Dorenbos, V. Zwiller, R. H. Hadfield, and G. S. Buller, "Kilometer-range, high resolution depth imaging via 1560 nm wavelength single-photon detection," Opt. Express **21**, 8904 (2013).

15. J. M. Bioucas-Dias and M. A. T. Figueiredo, "A New TwIST: Two-Step Iterative Shrinkage/Thresholding Algorithms for Image Restoration," Image Process. IEEE Trans. **16**, 2992–3004 (2007).

16. Z. T. Harmany, R. F. Marcia, and R. M. Willett, "This is SPIRAL-TAP: Sparse Poisson Intensity Reconstruction ALgorithms—Theory and Practice," IEEE Trans. Image Process. **21**, 1084–1096 (2012).

17. S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," Found. Trends Mach. Learn. **3**, 1–122 (2011).

18. R. L. Siddon, "Fast calculation of the exact radiological path for a three-dimensional CT array," Med. Phys. **12**, 252–255 (1985).

19. P. C. Hansen, J. G. Nagy, and D. P. O'leary, *Deblurring Images: Matrices, Spectra, and Filtering* (Siam, 2006), Vol. 3.

20. A. Beck and M. Teboulle, "Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems," IEEE Trans. Image Process. **18**, 2419–2434 (2009).

21. B. Zhang, J. M. Fadili, and J. L. Starck, "Wavelets, ridgelets, and curvelets for poisson noise removal," IEEE Trans. Image Process. **17**, 1093–1108 (2008).

22. Z. Zhu, Y. Sun, J. White, Z. Chang, and S. Pang, "Signal retrieval with measurement system knowledge using variational generative model," arXiv Prepr. arXiv1909.04188 (2019).

23. D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," (2013).

24. W. Chen, D. Kolditz, M. Beister, R. Bohle, and W. A. Kalender, "Fast on-site Monte Carlo tool for dose calculations in CT applications," Med. Phys. **39**, 2985–2996 (2012).

25. D. Kolditz, Y. Kyriakou, and W. A. Kalender, "Volume-of-interest (VOI) imaging in C-arm flat-detector CT for high image quality at reduced dose," Med. Phys. **37**, 2719–2730 (2010).

26. M. Hadwiger, L. Fritz, C. Rezk-Salama, T. Höllt, G. Geier, and T. Pabel, "Interactive volume exploration for feature detection and quantification in industrial CT data," IEEE Trans. Vis. Comput. Graph. **14**, 1507–1514 (2008).

27. G. Tisson, P. Scheunders, and D. Van Dyck, "3D region of interest x-ray CT for geometric magnification from multiresolution acquisitions," in *2004 2nd IEEE International Symposium on Biomedical Imaging: Macro to Nano (IEEE Cat No. 04EX821)* (IEEE, 2004), Vol. 2, pp. 567–570.

28. N. Mail, D. J. Moseley, J. H. Siewerdsen, and D. A. Jaffray, "The influence of bowtie filtration on cone-beam CT image quality.," Med. Phys. **36**, 22–32 (2009).

29. Z. Zhu, H.-H. Huang, and S. Pang, "Photon allocation strategy in region-of-interest tomographic imaging," IEEE Trans. Comput. Imaging 1–1 (2019).

30. D. Zhang and T. Z. Huang, "Generalized Tikhonov regularization method for large-scale linear inverse problems," J. Comput. Anal. Appl. **15**, 1317–1331 (2013).

31. K. W. DeLong, D. N. Fittinghoff, and R. Trebino, "Practical issues in ultrashort-laser-pulse measurement using frequency-resolved optical gating," IEEE J. Quantum Electron. **32**, 1253–1264 (1996).

32. D. J. Kennedy and S. T. Manson, "Photoionization of the noble gases: Cross sections and angular distributions," Phys. Rev. A **5**, 227–247 (1972).

33. F. T. Communication, "Coulomb – laser coupling in laser-assisted photoionization and molecular tomography Coulomb – laser coupling in laser-assisted photoionization and molecular tomography," (2007).

34. M. Ivanov and O. Smirnova, "How accurate is the attosecond streak camera?," Phys. Rev. Lett. **107**, 1–5 (2011).

35. J. White and Z. Chang, "Attosecond streaking phase retrieval with neural network," Opt. Express **27**, 4799 (2019).

36. K. Sohn, H. Lee, and X. Yan, "Learning Structured Output Representation using Deep Conditional Generative Models," Adv. Neural Inf. Process. Syst. 3483–3491 (2015).