ADVANCING SUPER RESOLUTION MICROSCOPY FOR QUANTITATIVE IN-VIVO IMAGING OF CHROMATIN NANODOMAINS

by

Clayton Seitz

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Physics

Indianapolis, Indiana

December 2024

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Gautam Vemuri, Chair

Department of Physics

Dr. Jing Liu

Department of Physics

Dr. Ruihua Cheng

Department of Physics

Dr. Stephen Wassall

Department of Physics

Dr. Horia Petrache

Department of Physics

Approved by:

TBD

I dedicate this thesis to those who have encouraged my pursuit of a doctoral degree.

To deal with a 14-dimensional space, visualize a 3-dimensional space and say 'fourteen' to yourself very loudly. Everyone does it

Geoffrey Hinton

Information is the resolution of uncertainty

Claude Shannon

ACKNOWLEDGMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS

m mass

v velocity

ABBREVIATIONS

abbr	abbreviation
bcf	billion cubic feet
BMOC	big man on campus

ABSTRACT

Single-molecule localization microscopy (SMLM) techniques, such as direct stochastic optical reconstruction microscopy (dSTORM), can be used to produce a pointillist representation of fluorescently-labeled biological structures at diffraction-unlimited precision. Direct STORM approaches leverage the deactivation of standard fluorescent tags, followed by spontaneous or photoinduced reactivation, allowing resolution of fluorophores at distances below the diffraction limit. This dissertation introduces single molecule localization microscopy and covers its application as discussed in the following papers:

BRD4 phosphorylation regulates structure of chromatin nanodomains [1] describes the role of the BRD4 phosphoswitch in the maintenance of chromatin nanodomains via super resolution microscopy and molecular dynamics simulation. We build on the notion that chromatin binding activity of BRD4 is regulated by phosphorylation by demonstrating that BRD4 phosphorylation regulated chromatin packing and mobility in mammalian nuclei.

Uncertainty-aware localization microscopy by variational diffusion [2] describes a novel algorithm that leverages a diffusion model in order to model a posterior distribution on super-resolution localization microscopy images. Fast extraction of physically relevant information from images using deep neural networks has led to significant advances in fluorescence microscopy and its application to the study of biological systems. For example, the application of deep networks for kernel density (KD) estimation in single molecule localization microscopy (SMLM) has accelerated super-resolution imaging of densely-labeled structures in the cell. However, simple and interpretable uncertainty quantification is lacking in these applications, and remains a necessary modeling component in high-risk research. We propose a generative modeling framework for KD estimation in SMLM based on variational diffusion. This approach allows us to probe the structure of the posterior on KD estimates, creating an additional avenue toward quality control. We demonstrate that data augmentation with traditional SMLM architectures followed by a diffusion process permits simultaneous high-fidelity super-resolution with uncertainty estimation of regressed KDEs.

1. Single molecule localization microscopy

1.1 Introduction

In the quest to understand cellular function, biologists aim to directly observe the processes enabling cells to maintain homeostasis and respond dynamically to internal and environmental cues at the molecular level. Super-resolution (SR) microscopy techniques have emerged as a pathway to this aim, surpassing the classical Abbe diffraction limit of optical resolution: $\lambda/2\text{NA}$ where λ is the emission wavelength and NA is the numerical aperture of an objective lens. Fluorescence microscopy techniques continually push the resolution boundary towards nanometer scales, facilitating imaging of cellular structures with a level of detail previously achievable only with electron microscopy (EM). Concurrently, SR techniques retain optical microscopy advantages in biological experiments, including sample preservation, imaging flexibility, and target specificity. SR enables extraction of quantitative information on spatial distributions and often absolute numbers of proteins, nucleic acids, or other macromolecules within subcellular compartments. [1]

Many SR methods are based on wide-field (WF), total internal reflection fluorescence (TIRF) or confocal microscope setups and fundamentally differ in how fluorescently labeled samples are excited and how the emitted photons are detected. Here, I focus on single-molecule localization microscopy (SMLM) techniques a class of SR diffraction-unlimited SR methods which leverage fluorescence intermittency to resolve fluorophores in the sample whos spatially overlapping point spread functions would otherwise render them unresolvable at the detector. SMLM approaches, such as direct-STORM (dSTORM) have become quite popular because they can be implemented at low cost on conventional, camera-based, wide-field setups, shifting the complexity to biological sample preparation and image post processing. Common strategies for the temporal separation of molecules involve transient intramolecular rearrangements to switch from dark to fluorescent states or the exploitation of non-emitting molecular radicals. For example, in dSTORM, rhodamine derivatives can undergo intersystem crossing to a triplet state, which can be reduced by thiols to form a dark radical species. The dark state can then be quenched by oxidative processes, driving the fluorophore back to its ground state.

In SMLM applications, we seek the position and intensity of isolated fluorophores as well as to estimate the accuracy and precision of these parameters. Accuracy is a measure of the systematic error or bias, and precision is a measure of the statistical error of an estimator. To generate super-resolution images using SMLM, single emitters are located, and using the mosaic of their found positions, we produce a kernel density estimate (KDE). Such KDEs are often Gaussian, and are used to generate the final super-resolution images. The width of one such placed Gaussian function, σ is given by the precision of the fluorophore position localization. Therefore, in SMLM, it is necessary to both find the parameters and estimate their precision. Reported values are in the range of 2070 nanometers. In the following section, we derive a fundamental statistical description of fluorophore detection in SMLM, which is compatible with a coherent state of the quantized electromagnetic field. This description is necessarily simplified - background rates of light detection may vary across the field of view, and the fluorophore emission rate of chemically identical fluorophores can vary owing to effects such as uneven illumination profile, dipole orientation or different optical path lengths.

The central objective of SMLM is to infer a set of molecular coordinates $\theta = (\theta_u, \theta_v)$ from measured low resolution images $\mathbf{x}$. The likelihood on a particular pixel $p(\mathbf{x}_k|\theta)$ is taken to be a convolution of Poisson and Gaussian distributions, due to shot noise $p(s_k) = \text{Poisson}(\omega_k)$ and sensor readout noise $p(\zeta_k) = \mathcal{N}(o_k, w_k^2)$

$$p(\mathbf{x}_k|\theta) = A\sum_{q=0}^{\infty}\frac{1}{q!}\mathrm{e}^{-\omega_k}\omega_k^q\frac{1}{\sqrt{2\pi}w_k}\mathrm{e}^{-\frac{(\mathbf{x}_k-g_kq-o_k)^2}{2w_k^2}} \approx \text{Poisson}(\omega_k') \tag{1.1}$$

where A is some normalization constant. For the sake of generality, we include a per-pixel gain factor g_k, which is often unity. Sampling from $p(\mathbf{x}_k|\theta)$ is trivial; however, for computation of a lower bound on uncertainty in θ, the summation in (1) can be difficult to work with. Therefore, we choose to use a Poisson approximation for simplification, valid under a range of experimental conditions (Huang2013). After subtraction of a known offset o_k of the pixel array, which can be easily measured, we have $\omega_k' = \omega_k + w_k^2$. The expectation of the Poisson process ω_k at each pixel of the image must then be computed from the optical impulse response $O(u, v)$, which is taken to be an isotropic Gaussian in two-dimensions.

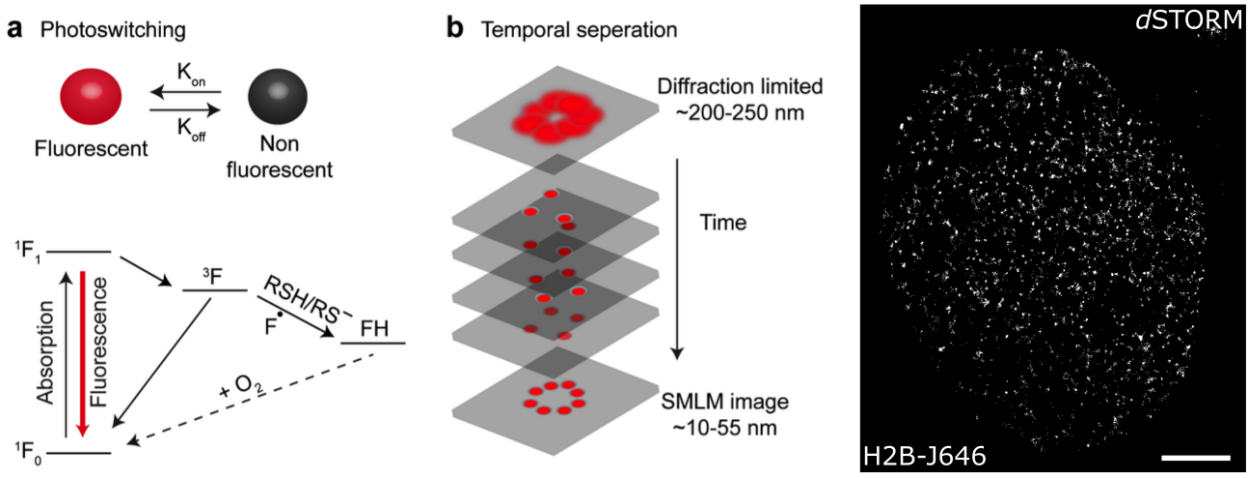

Figure 1.1. Stochastic optical reconstruction microscopy (STORM). (A) Single molecules are resolved by separating their fluorescent emission in time, using fluorophores with multiple photophysical states (B) Example super-resolution image of H2B protein in a living Hela cell nucleus at 37C, 5 percent CO2. Image reconstructed from 10^3 10ms frames. Scalebar 5um.

Consider an idealized scenario where an isolated fluorescent molecule exists in the image plane. For a particular pixel k of width δ centered at $\xi_k = (u_k, v_k)$, we define, along the first image dimension

$$\Delta E_u := \int_{u_k-\delta/2}^{u_k+\delta/2} O(u;\theta_u)du = \frac{1}{2}\left(\text{erf}\left(\frac{u_k+\frac{\delta}{2}-\theta_u}{\sqrt{2}\sigma_{\mathbf{x}}}\right) - \text{erf}\left(\frac{u_k-\frac{\delta}{2}-\theta_u}{\sqrt{2}\sigma_{\mathbf{x}}}\right)\right) \tag{1.2}$$

The expected value at each pixel is then $\omega_k \propto \Delta E_u(u_k, \theta_u, \sigma_{\mathbf{x}})\Delta E_v(v_k, \theta_v, \sigma_{\mathbf{x}})$. Using this, sampling from the convolution distribution in (1) is can be carried out by $\mathbf{x}_k = s_k + \zeta_k$ for $s_k \sim \text{Poisson}(\omega_k), \eta_k \sim \mathcal{N}(o_k, w_k^2)$.

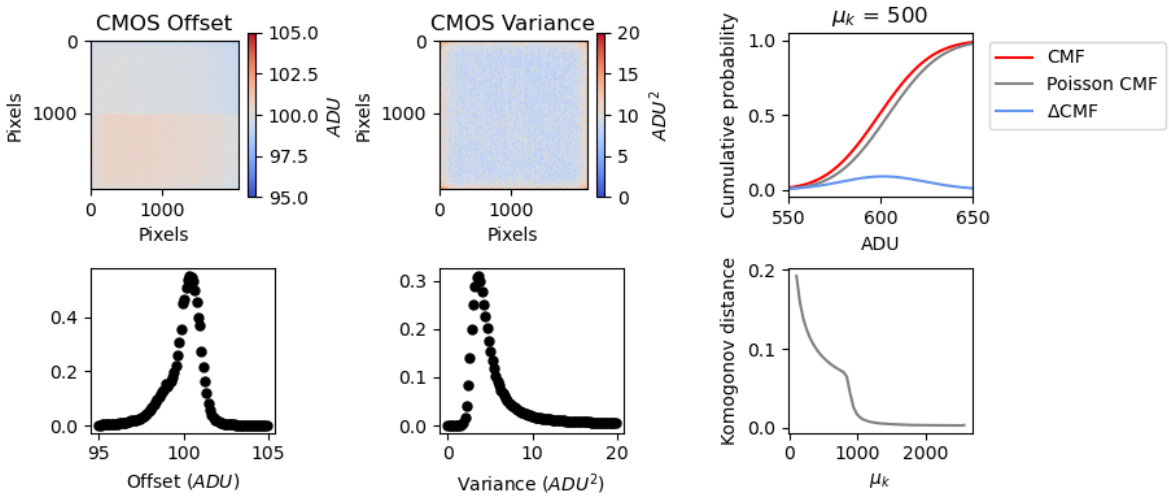

Figure 1.2. Noise model for CMOS cameras used for MLE. (left)) CMOS offset for zero incident photons (middle) CMOS variance for zero incident photons (upper right) Cumulative mass function for the convolution distribution and its Poisson approximation for rate parameter $\mu_k = 500$ counts (lower right) Komogonov distance measured as a function of rate parameter μ_k

1.1.1 The definition of resolution in SMLM

The distribution of a particular biomolecule in the cell can be described as a probability density over a two-dimensional space, casting super-resolution as a density estimation problem. Intuitively, the spatial resolution of SMLM images then increases as we draw more samples from this density - a concept which is made mathematically precise by the so-called Fourier ring correlation or FRC. Using FRC, one can compute image resolution as the spatial frequency at which a correlation function in the frequency domain drops below a threshold, typically taken to be 1/7 (See Supplement). According to this theory, reducing localization uncertainty while increasing the number of samples, results in an increase in image resolution (Nieuwenhuizen 2013). However, there remains a fundamental limit to the the minimal localization uncertainty which can be obtained.

$$\mathrm{FRC}(q) = \frac{\sum_{\vec{q}\in\mathrm{circle}} \tilde{f}_1(\vec{q})\tilde{f}_2(\vec{q})^*}{\sqrt{\sum_{\vec{q}\in\mathrm{circle}}|f_1(\vec{q})|^2}\sqrt{\sum_{\vec{q}\in\mathrm{circle}}|f_2(\vec{q})|^2}}$$

Localization uncertainty, typically the RMSE of a maximum likelihood or similar statistical estimator, is bounded from below by the inverse of the Fisher information matrix, known as the Cramer-Rao lower bound (Chao 2016). Localization uncertainties in sparse conditions are often tens of nanometers, although recent work on integration of Bayesian priors with modulation enhanced SMLM (meSMLM) or structured illumination with MINFLUX, has reduced spatial resolution below to a few nanometers (Kalisvaart 2022, Gwosh 2020). Nevertheless, managing the increase in localization uncertainty at high labeling density remains a major bottleneck to SMLM. Static uncertainty due to molecular crowding can be partially amelioriated by using pairwise or higher-order temporal correlations within a pixel neighborhood, known as stochastic optical fluctuation imaging or SOFI (Dertinger 2009). Other approaches such as stimulated emission and depletion (STED) imaging bring control over the photophysical state of a chosen subset of the sample, yet the need for laser scanning prevents widespread application in live-cell studies. The spatial resolution and relative simplicity of SMLM techniques remains unmatched, inciting an effort to increase the resolution of SMLM techniques and explore avenues towards time resolved SMLM.

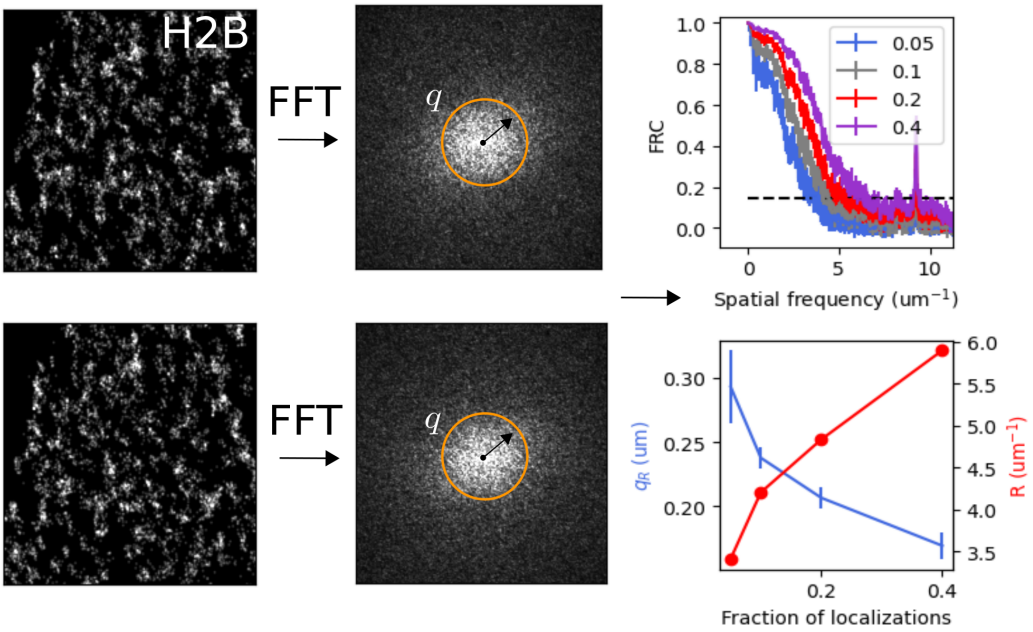

Figure 1.3. Noise model for CMOS cameras used for MLE. (left)) CMOS offset for zero incident photons (middle) CMOS variance for zero incident photons (upper right) Cumulative mass function for the convolution distribution and its Poisson approximation for rate parameter $\mu_k = 500$ counts (lower right) Komogonov distance measured as a function of rate parameter μ_k

2. Bromodomain protein 4 and chromatin organization

3. Uncertainty-aware localization microscopy by variational diffusion

In single molecule localization, uncertainty quantification is a critical aspect of parameter estimation. Uncertainty can be defined in either the frequentist or the Bayesian statistical paradigm. These primarily differ in their fundamental definition of probability. The frequentist defines probability in terms of the frequency of observed events and examines the empirical distribution of these samples to express uncertainties. The Bayesian would express uncertainty *apriori* based on, say, domain knowledge. Frequentist methods typically estimate uncertainties under a *fixed* model $\mathcal{M}$ using techniques such as a Laplace approximation around the maximum likelihood estimate (MLE). For example, a frequentist might compute

$$\mathcal{L}(D|\mathcal{M}) = \prod_{\mathrm{i}=1}^{N} P(D_{\mathrm{i}}|\mathcal{M})$$

for a dataset of size N, where the model M is held fixed. A frequentist might also make use of the Cramér-Rao lower bound (CRLB), which provides theoretical bounds on the uncertainty achievable by any unbiased estimator.

In contrast, Bayesian approaches to uncertainty quantification in single molecule localization often involve parameter exploration techniques such as Markov Chain Monte Carlo (MCMC). MCMC methods sample from the posterior distribution over the parameters, allowing for comprehensive exploration of parameter space and estimation of uncertainty.

$$P(\mathcal{M}|D) \propto P(D|\mathcal{M})P(\mathcal{M})$$

which can also be optimized by maximum aposteriori estimation (MAP). Additionally, Bayesians may utilize inequalities analagous to the CRLB such as the van Trees inequality, which provides bounds on the expected error in parameter estimation. In the context of single molecule localization, where only a single measurement is often feasible, the treatment of uncertainty demands a Bayesian perspective. We are unable to consider the data that we could have gotten; rather, we must express our uncertainties conditioned on the data that we do have. However, Bayesian approaches are often avoided in this domain, due to limited

computational resources. Diffusion models are a promising direction to speed up and simplify Bayesian inference.

REFERENCES

[1] H. Kong *et al.*, "A generalized laplacian of gaussian filter for blob detection and its applications," *IEEE Transactions on Cybernetics*, vol. 43, pp. 1719–1733, 2013.