

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

by

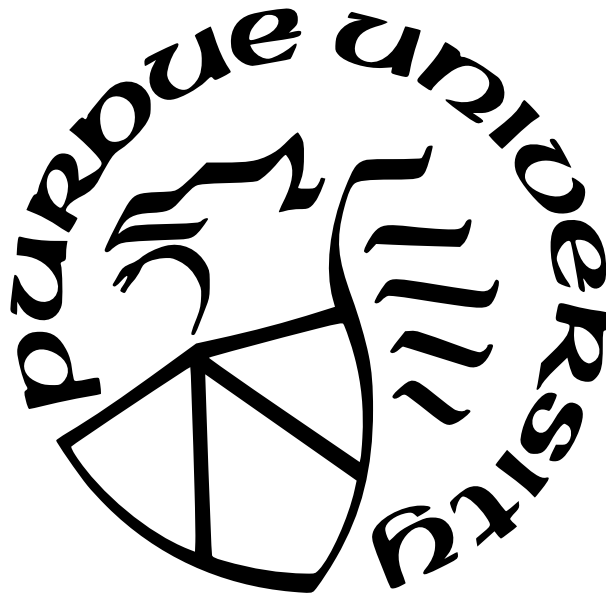
Clayton Seitz

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	7
LIST OF FIGURES	8
LIST OF SYMBOLS	9
ABSTRACT	11
1 Single molecule localization microscopy	12
1.1 Introduction	12
1.2 The Image Likelihood	14
1.2.1 The definition of resolution in SMLM	15
2 Bromodomain protein 4 and chromatin organization	18
2.1 Introduction	18
3 Uncertainty-aware localization microscopy by variational diffusion	19
3.1 Introduction	19
3.2 Background	21
3.2.1 Gaussian kernel density estimation	21
3.3 Uncertainty-Aware Localization Microscopy by Variational Diffusion	22
3.3.1 Problem Statement	22
3.3.2 Variational Diffusion	23
3.4 Experiments	25
3.5 Conclusion	27
3.6 Broader Impact	27
3.7 Appendix	28
3.7.1 Sampling	28
3.7.2 Derivation of the variational bound	28
3.7.3 Metropolis-Hastings MCMC	30
3.7.4 Cramer-Rao Lower Bound	31

3.7.5	Neural Networks ψ, ϕ	31
4	Enhanced SMLM with single photon counting cameras	34
4.0.1	Introduction	34
4.0.2	Basic Scheme	35
4.0.3	Results	39
4.0.4	Discussion	41

LIST OF TABLES

LIST OF FIGURES

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.	13
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	16
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	17
3.1	Estimation of the marginal variances $\sqrt{\text{Var}(\mathbf{y}_k)}$ for an isolated fluorescent emitter. MCMC sampling is carried out using the low-resolution image (left) to estimate the posterior $p(\theta \mathbf{x})$ (middle) which is sampled $\theta \sim p(\theta \mathbf{x})$ and combined with (3) to estimate marginal variances (right)	21
3.2	Non cherry-picked estimation of marginal variances. A low-resolution image \mathbf{x} (left column) is transformed by ϕ to produce a KDE estimate $\hat{\mathbf{y}}$ (middle column) and a DDPM ψ computes a map of marginal variances (right column)	26
3.3	Localization errors of the trained model ϕ . The Cramer-Rao lower bound is shown in red, computing by taking the diagonal elements of $I^{-1}(\theta)$	32
4.1	Single photon counting with a SPAD array (a) Conventional widefield microscopy with integrated SPAD array (b) Single photon imaging scheme using 1us exposures containing a picosecond laser pulse (c) Sum of photon counts over a 5x5 region of interest (ROI), taken with $N_{\text{frames}} = 5 \times 10^5$	36
4.2	Single and multi-emitter localization error on sums of photon counts. (left) Localization uncertainty for simulated data for different values of N , plotted with respect to the Cramer-Rao lower bound, shown in dashed gray. (right) Multi-emitter localization by MCMC sampling for $N = 3$, colors indicate a cluster of samples i.e., a single localization. All data was generated with a background rate $\langle \mathbf{n}_{\text{background}} \rangle = \lambda N_{\text{frames}}/d^2$ per pixel. Scalebar 360nm	37
4.3	Posteriors on the number of fluorescent emitters N and localization for $N^* = 1$ (left) and $N^* = 2$ (right) quantum dots. Scalebars 360nm	39

LIST OF SYMBOLS

\mathbf{x}	An image at base resolution
\mathbf{y}	An image at higher resolution
δ	Pixel lateral width
k	Pixel index
θ	A parameter
u, v	Cartesian coordinates in two-dimensions
g_k	Pixel-wise gain
o_k	Pixel-wise offset
w_k	Pixel-wise readout noise standard deviation
μ_k	Pixel-wise expected value
s_k	Pixel-wise measured signal
ξ_k	Pixel-wise measured readout noise
O	Point spread function
erf	Error function
$\sigma_{\mathbf{x}}$	Gaussian PSF width
$\sigma_{\mathbf{y}}$	Kernel width for kernel density estimate
ϵ	An image of pure Gaussian noise
β	Diffusion model noise variance
SNR	Diffusion model signal to noise ratio
ψ	Diffusion model parameters
ϕ	Augmentation network parameters
\mathcal{L}	An objective function
D_{KL}	KL-divergence
p	A discrete or continuous probability distribution
\mathbb{E}_p	Expectation with respect to a distribution p
I	Fisher information matrix
λ	Expected number of background counts per frame
d	Lateral dimension of a region of interest

ζ	Photon detection probability
τ	Delay time
$g^{(2)}(\tau)$	Second order coherence function
B	Expected number of signal-background coincidences per frame
$G^{(2)}(m)$	Measured number of signal-signal coincidences at lag time m
N_{frames}	Number of frames
N	Number of active fluorescent emitters
N^*	Maximum a posteriori estimate of number of active emitters
n	Number of photon counts in a frame
ℓ	Log-likelihood
\hat{a}	Ladder operator
\hat{E}	Electric field operator
ρ	Density matrix or number density for molecular dynamics
$L(r)$	Besag's L-function
$K(r)$	Ripley's K-function
$G(r)$	Nearest neighbor distribution function
λ	(Point Pattern Analysis) Intensity of a point process
γ	Friction tensor
ξ	A delta-correlated Gaussian noise
k_B	Boltzmann's constant
T	Temperature
U	Potential energy
D	Diffusion coefficient
ϵ	(Molecular Dynamics) Energy
r_0	Harmonic bond equilibrium length
R_0	Binder potential equilibrium length
α	Anomalous diffusion exponent

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. [Kong2013]

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.

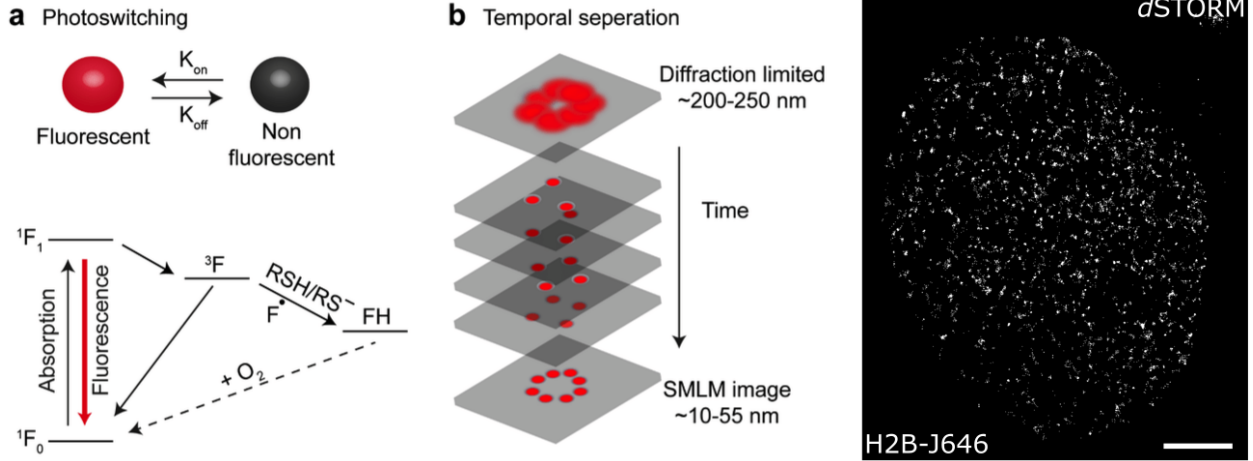


Figure 1.1. Stochastic optical reconstruction microscopy (STORM).

(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 5μm.

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.

1.2 The Image Likelihood

It is common to describe the optical impulse response of a microscope as a two-dimensional isotropic Gaussian [Zhang2007]. This is an approximation to the more rigorous diffraction models given by [Richards1959, Gibson1989]. Over a continuous domain, the impulse response reads

$$O(u, v) = \frac{1}{2\pi\sigma_{\mathbf{x}}^2} e^{-\frac{(u-\theta_u)^2 + (v-\theta_v)^2}{2\sigma_{\mathbf{x}}^2}}$$

The above expression can be interpreted as a probability distribution over locations where a photon can be detected. Therefore, for discrete detectors, we discretize this expression by integrating over pixels. The number of photon arrivals will follow Poisson statistics, with expected value

$$\mu_k = i_0 \left(\int_{u_k - \delta/2}^{u_k + \delta/2} O(u; \theta_u) du \right) \left(\int_{v_k - \delta/2}^{v_k + \delta/2} O(v; \theta_v) dv \right)$$

The scalar quantity i_0 represents the amplitude of the signal, which is proportional the quantum efficiency of a pixel η , the duration of exposure, Δ , and the number of photons emitter by a fluorescent molecule N_0 . With no loss of generality, $\Delta = \eta = 1$ and there is a single free parameter N_0 . A simple change of variables $u' = u - \theta_u$ and $v' = v - \theta_v$ gives

$$\mu_k = i_0 \left(\int_{u_k - \delta/2 - \theta_u}^{u_k + \delta/2 - \theta_u} O(u) du \right) \left(\int_{v_k - \delta/2 - \theta_v}^{v_k + \delta/2 - \theta_v} O(v) dv \right)$$

Terms in parentheses are simply integrals of Gaussian functions and can be evaluated analytically. I define them as Γ_u, Γ_v , respectively. For example,

$$\begin{aligned} \Gamma_u &:= \int_{u_k - \delta/2 - \theta_u}^{u_k + \delta/2 - \theta_u} O(u) du \\ &= \int_0^{u_k + \delta/2 - \theta_u} O(u) du - \int_0^{u_k - \delta/2 - \theta_u} O(u) du \\ &= \frac{1}{2} \left(\operatorname{erf} \left(\frac{u_k + \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) - \operatorname{erf} \left(\frac{u_k - \frac{\delta}{2} - \theta_i}{\sqrt{2}\sigma_{\mathbf{x}}} \right) \right) \end{aligned}$$

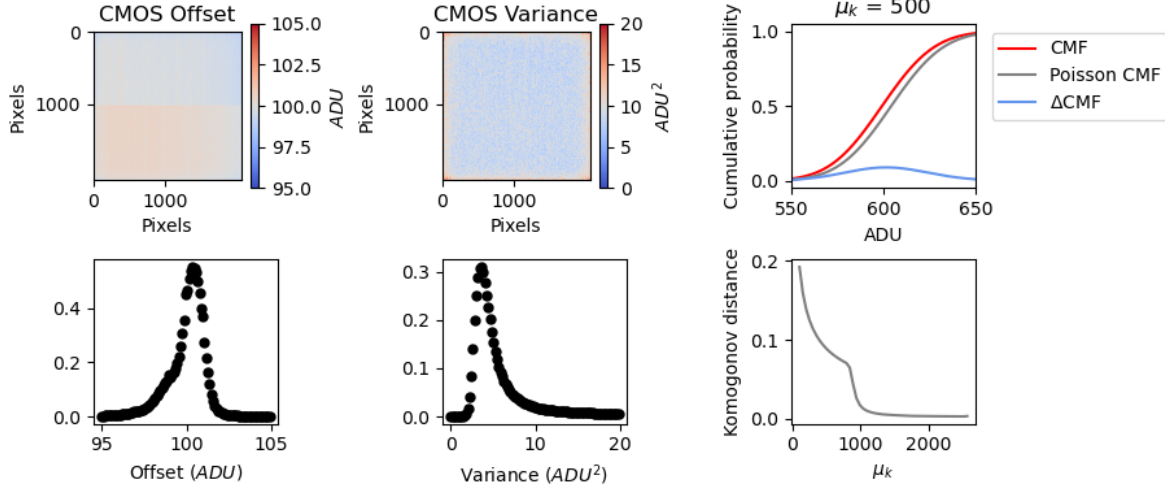


Figure 1.2. Noise model for CMOS cameras used for MLE. (left) CMOS offset for zero incident photons (middle) CMOS variance for zero incident photons (upper right) Cumulative mass function for the convolution distribution and its Poisson approximation for rate parameter $\mu_k = 500$ counts (lower right)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

where we have used the common definition $\text{erf}(z) = \frac{2}{\sqrt{\pi}} \int_0^z e^{-t^2} dt$. Recall 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} . In general, 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 $p(s_k) = \text{Poisson}(\mu_k)$ and sensor readout noise $p(\xi_k) = \mathcal{N}(o_k, w_k^2)$

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

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 $\mu'_k = \mu_k + w_k^2$.

Under the Poisson approximation in the likelihood, the model negative log-likelihood is

$$\ell(\mathbf{x}|\theta) = -\log \prod_k \frac{e^{-(\mu'_k)} (\mu'_k)^{n_k}}{n_k!} = \sum_k \log n_k! + \mu'_k - n_k \log (\mu'_k) \quad (1.2)$$

Localization then proceeds by minimization of $\ell(\mathbf{x}|\theta)$.

1.2.1 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.

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

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 MIN-FLUX, 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 ameliorated by using pairwise or higher-order temporal correlations within a pixel neighborhood, known as stochastic optical fluctuation imaging or SOFI (Dertinger

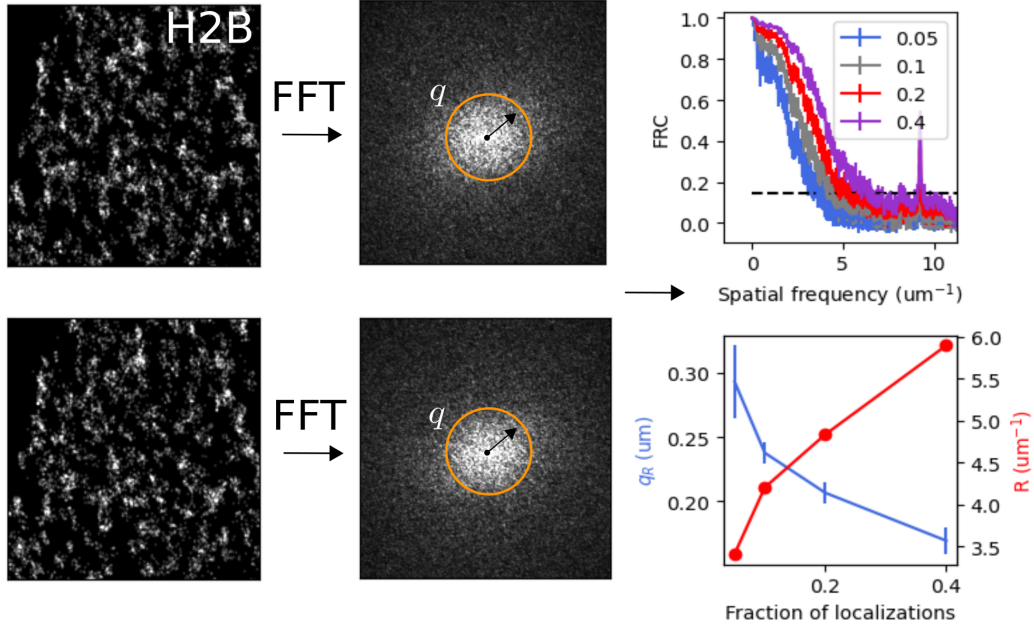


Figure 1.3. Noise model for CMOS cameras used for MLE. (left) CMOS offset for zero incident photons (middle) CMOS variance for zero incident photons (upper right) Cumulative mass function for the convolution distribution and its Poisson approximation for rate parameter $\mu_k = 500$ counts (lower right)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

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.

2. Bromodomain protein 4 and chromatin organization

The interplay between chromatin structure and phase-separating proteins is an emerging topic in cell biology with implications for understanding disease states. Here, we investigate the functional relationship between bromodomain protein 4 (BRD4) and chromatin architecture. By combining molecular dynamics simulations with live-cell imaging, we demonstrate that BRD4, when phosphorylated at specific N-terminus sites, significantly impacts nucleosome nanodomain (NN) organization and dynamics. Our findings reveal that enhanced chromatin binding activity of BRD4 condenses NNs, while both loss or gain of BRD4 chromatin binding reduced diffusion of single nucleosomes, suggesting a role for BRD4 in the regulation of nanoscale chromatin architecture and the chromatin microenvironment. These observations shed light on the nuanced regulation of chromatin structure by BRD4, offering insights into its role in maintaining the nuclear architecture and transcriptional activity.

2.1 Introduction

3. Uncertainty-aware localization microscopy by variational diffusion

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.

3.1 Introduction

Deep models have attracted tremendous attention from researchers in the natural sciences, with several foundational applications arising in microscopy [Weigert2018, Falk2019]. Recently, the application of deep image translation in single-molecule localization microscopy (SMLM) has received considerable interest. SMLM techniques are a mainstay of fluorescence microscopy, which localize “blinking” fluorescent molecules to produce a pointillist representation of biomolecules in the cell at diffraction-unlimited precision [Rust2006, Betzig2006]. Recently, the use of deep models to perform localization has been proposed as an alternative to traditional localization algorithms, in order to increase imaging speed and labeling density. In previous applications of deep models to localization microscopy, super-resolution images have been recovered from a sparse set of localizations with conditional generative adversarial networks [Ouyang2018] or localization itself can be performed using traditional convolutional networks [Nehme2020, Speiser2021]. In this paper, we perform localiza-

tion indirectly by predicting kernel density (KD) estimates of a population of fluorescent molecules using a deep model.

Kernel density estimation in SMLM is necessarily performed using a single low-resolution image, and thus common measures of model performance are based on localization errors computed over ensembles of simulated images. Unfortunately, this choice precludes computation of uncertainty at test time under a fixed model. Bayesian probability theory is therefore an attractive alternative, which offers us mathematically grounded tools to reason about uncertainty.

We model a posterior on high-resolution KD estimates conditioned on a low-resolution image. Our approach is based on a type of score based generative model [Song2021], referred to as a denoising diffusion probabilistic model (DDPM) in the literature [Ho2020, Song2021]. We find that this technique is complementary to relevant existing approaches to uncertainty estimation, which would primarily address epistemic sources of uncertainty, using techniques such as ensembling [Lakshminarayanan2017] or Monte Carlo dropout [Gal2022]. The approach is inspired by recent variational perspectives on diffusion [Dirmeier2023, Ribeiro2024, Kingma2021, Kingma2023]. Such techniques provide a mechanism for scalable variational inference, which can be trained using a variational bound written in terms of the signal-to-noise ratio of the diffused data, and a simple noise estimation loss. Indeed, recent efforts have shown that the variational bound can be reparameterized to give several more conventional diffusion losses [Kingma2021, Kingma2023, Ribeiro2024].

In the remainder of this paper, we first introduce the likelihood of low-resolution images in localization microscopy, and show uncertainty quantification in a rudimentary example scenario. Then, we introduce KD estimation as an alternative to direct localization using low-resolution images, followed by demonstration of our variational diffusion model for measuring uncertainty KD estimation at scale.

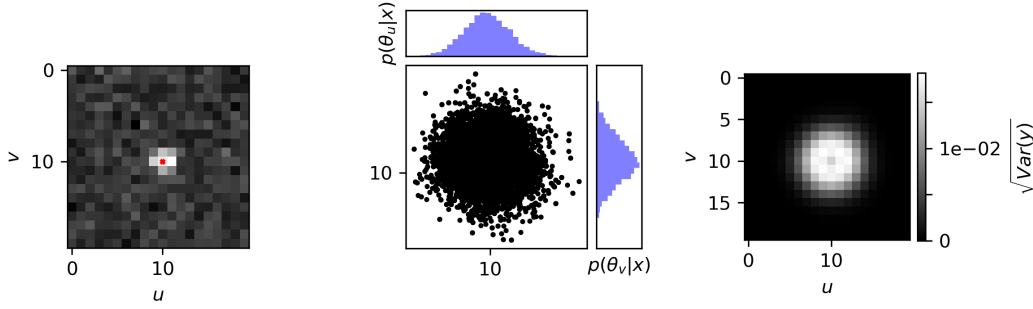


Figure 3.1. Estimation of the marginal variances $\sqrt{\text{Var}(\mathbf{y}_k)}$ for an isolated fluorescent emitter. MCMC sampling is carried out using the low-resolution image (left) to estimate the posterior $p(\theta|\mathbf{x})$ (middle) which is sampled $\theta \sim p(\theta|\mathbf{x})$ and combined with (3) to estimate marginal variances (right)

3.2 Background

3.2.1 Gaussian kernel density estimation

Direct optimization of the likelihood in (1) from observations \mathbf{x} alone is challenging when fluorescent emitters are dense within the field of view and fluorescent signals significantly overlap. However, convolutional neural networks (CNNs) have recently proven to be powerful tools fluorescence microscopy to extract parameters describing fluorescent emitters such as color, emitter orientation, z -coordinate, and background signal **Zhang2018**, **Kim2019**, **Zelger2018**. For localization tasks, CNNs typically employ upsampling layers to reconstruct Bernoulli probabilities of emitter occupancy [**Speiser2021**] or KD estimates with higher resolution than experimental measurements [**Nehme2020**]. We choose to use KD estimates in our model, denoted by \mathbf{y} , which are latent in the low-resolution data \mathbf{x} . KDEs are the most common data structure used in SMLM, and can be easily generated from molecular coordinates, alongside observations \mathbf{x} .

Similar to the generative process on low resolution images \mathbf{x} , we generate KDEs \mathbf{y} by repurposing the generative model (1) on an unsampled image without noise. In other words, we cast Gaussian KD estimation as a noiseless image generation process on the domain of

\mathbf{y} . Under a fixed configuration of N particles θ , the value of a non-normalized KDE pixel \mathbf{y}_k is given by

$$\mathbf{y}_k(\theta) = \sum_{n=1}^N \Delta E_u(u_k, \theta_u, \sigma_{\mathbf{y}}) \Delta E_v(v_k, \theta_v, \sigma_{\mathbf{y}}) \quad (3.1)$$

where the hyperparameter $\sigma_{\mathbf{y}}$ is a Gaussian kernel width.

3.3 Uncertainty-Aware Localization Microscopy by Variational Diffusion

We now consider more realistic datasets $(\mathbf{x}_i, \mathbf{y}_{0,i}, \hat{\mathbf{y}}_i)_{i=1}^N$ of observed images \mathbf{x}_i true KD images $\mathbf{y}_{0,i}$, and augmented low-resolution inputs $\hat{\mathbf{y}}_i = \phi(\mathbf{x}_i)$, where ϕ is a CNN. Observations \mathbf{x}_i are simulated under the convolution distribution (1) and KDEs are generated by (4).

3.3.1 Problem Statement

Kernel density estimates produced by the traditional deep architectures for localization microscopy produce strong results, but lack uncertainty quantification. Unfortunately, the posterior $p(\theta|\mathbf{x})$ has no known analytical form and can be difficult to compute at test time, since (i) molecules cannot be easily resolved and therefore θ is of unknown dimension and (ii) θ can be high dimensional and efficient exploration of the parameter space is challenging. The central goal of this paper is to instead model a conditional distribution on the latent \mathbf{y} : $p(\mathbf{y}|\mathbf{x})$, where \mathbf{y} is of known dimensionality. We choose to model $p(\mathbf{y}|\mathbf{x})$ with a diffusion model, given that the distribution $p(\mathbf{y}|\mathbf{x})$ is expensive to compute, even if $p(\theta|\mathbf{x})$ were known.

Recent advances in generative modeling, particularly diffusion models [SohlDickstein2015, Ho2020, Song2021] present a unique opportunity to integrate uncertainty awareness into the localization microscopy toolkit. However, sampling from diffusion models can be computationally expensive, given that generation amounts to solving a complex stochastic differential equation, effectively mapping a simple base distribution to the complex data distribution. The solution of such equations requires numerical integration with very small step sizes, resulting in thousands of neural network evaluations [Saharia2021, Vahdat2021]. For conditional generation tasks in high-risk applications, generation complexity is further exacerbated by the need for the highest level of detail in generated samples. Therefore, we

propose that sampling is preceded by an augmentation network ϕ , which in essence generates an initial estimate to guide the diffusion process. Reasoning for this choice in our application is two-fold:

Synthesis Speed. By training the augmentation network ϕ to obtain an approximate estimate of \mathbf{y}_0 , we can reduce the number of iterations, since the diffusion model only needs to model the remaining mismatch, resulting in a less complex model from which sampling becomes easier. Speed is critical in SMLM applications, which can produce thousands of images in a single experiment.

Sample Fidelity. Since Langevin dynamics will often be initialized in low-density regions of the data distribution, inaccurate score estimation in these regions will negatively affect the sampling process. Moreover, mixing can be difficult because of the need of traversing low density regions to transition between modes of the distribution [Song2019].

3.3.2 Variational Diffusion

Diffusion models [SohlDickstein2015, Ho2020, Song2021] are a class of generative models originally inspired by nonequilibrium statistical physics, which slowly destroy structure in a data distribution via a fixed Markov chain referred to as the *forward process*. In the present context, we leverage the variational interpretation of this model class [Kingma2021, Kingma2023] to approximate the posterior $p(\mathbf{y}|\mathbf{x})$.

Diffusion Model. We use a forward process which gradually adds Gaussian noise to the latent \mathbf{y}_0 in discrete time, according to a variance schedule β_t :

$$q(\mathbf{y}_T|\mathbf{y}_0) = \prod_{t=1}^T q(\mathbf{y}_t|\mathbf{y}_{t-1}) \quad q(\mathbf{y}_t|\mathbf{y}_{t-1}) = \mathcal{N}\left(\sqrt{1-\beta_t}\mathbf{y}_{t-1}, \beta_t I\right) \quad (3.2)$$

An important property of the forward process is that it admits sampling \mathbf{y}_t at an arbitrary timestep t in closed form [Ho2020]. Using the notation $\alpha_t := 1 - \beta_t$ and $\gamma_t := \prod_{s=1}^t \alpha_s$, we have $q(\mathbf{y}_t|\mathbf{y}_0) = \mathcal{N}\left(\sqrt{\gamma_t}\mathbf{y}_0, (1 - \gamma_t)I\right)$ or $\mathbf{y}_t = \sqrt{\gamma_t}\mathbf{y}_0 + \sqrt{1 - \gamma_t}\epsilon$ for $\epsilon \sim \mathcal{N}(0, I)$. The signal to noise ratio (SNR) as defined in [Kingma2023], at a time step t reads $\text{SNR}_t = \gamma_t/(1 - \gamma_t)$.

The usual procedure is then to learn a parametric representation of the *reverse process*, and therefore generate samples of the latent \mathbf{y}_0 from $p(\mathbf{y}_0|\mathbf{x})$. Formally, $p_\psi(\mathbf{y}_0|\mathbf{x}) = \int p_\psi(\mathbf{y}_{0:T}|\mathbf{x}) d\mathbf{y}_{1:T}$ where \mathbf{y}_t is a latent representation with the same dimensionality of the data and $p_\psi(\mathbf{y}_{0:T}|\mathbf{x})$ is a Markov process, starting from a noise sample $p_\psi(\mathbf{y}_T) = \mathcal{N}(0, I)$. Writing this Markov process gives

$$p_\psi(\mathbf{y}_{0:T}|\mathbf{x}) = p_\psi(\mathbf{y}_T) \prod_{t=1}^T p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}) \quad p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x}) = \mathcal{N}(\mu_\psi(\mathbf{y}_t, \gamma_t), \beta_t I) \quad (3.3)$$

where we reuse the variance schedule of the forward process [Ho2020]. From (5) it can be seen that the learnable parameter in the reverse process is the expectation of the transition μ_ψ where ψ is a neural network.

Learning the reverse process can be approached by either regressing noise ϵ from the forward process, or the true latent \mathbf{y}_0 , as there is a deterministic relationship between them. We adopt the former for consistency with other work, and define ψ as a neural denoising function which regresses the noise ϵ from a noisy \mathbf{y}_t . A relation between the noise estimate ϵ_ψ and μ_ψ is given in the Appendix, which gives an intuition for sampling. The proposed sampling scheme is depicted in (Figure 3).

Variational Objective. Following [Kingma2021], we interpret the reverse process as a hierarchical generative model that samples a sequence of latents \mathbf{y}_t , with time running backward. Training of the model is achieved through the variational bound

$$-\log p(\mathbf{y}_0) \leq -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \left(\frac{p_\psi(\mathbf{y}_{1:T}, \mathbf{y}_0)}{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \right) \quad (3.4)$$

$$= D_{KL}(q(\mathbf{y}_T|\mathbf{y}_0)||p(\mathbf{y}_T)) + \mathbb{E}_{q(\mathbf{y}_1|\mathbf{y}_0)} \log p(\mathbf{y}_0|\mathbf{y}_1) + \mathcal{L}_\psi \quad (3.5)$$

where we have omitted conditioning on the low-resolution \mathbf{x} to simplify the notation. Note that, this is similar to a hierarchical VAE, but in a diffusion model $q(\mathbf{y}_{1:T}|\mathbf{y}_0)$ is fixed by the forward process rather than learned. The so-called diffusion loss \mathcal{L}_ψ is shown in Appendix A, and is the term of interest as the first two terms do not contribute meaningfully to the loss [Ho2020]. Furthermore, it has become standard to use simplified forms of \mathcal{L}_ψ , such as

a noise estimation loss, as this has shown superior performance. Importantly, \mathcal{L}_ψ is simply a reweighted variant of a family of diffusion objectives [Kingma2021, Kingma2023]. We use the following Monte Carlo estimate of \mathcal{L}_ψ , which demonstrates that the variational bound can be written in terms of the common noise-estimation loss

$$\mathcal{L}_\psi = \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I), t \sim U(1, T)} \left[\left(\frac{\text{SNR}_{t-1}}{\text{SNR}_t} - 1 \right) \|\epsilon - \epsilon_\psi\|_2^2 \right] \quad (3.6)$$

A full derivation of this objective is outlined in the Appendix. Note that SNR_t is monotonically decreasing with t , and thus $\frac{\text{SNR}_{t-1}}{\text{SNR}_t} = \frac{\gamma_{t-1}}{\gamma_t} \frac{1-\gamma_t}{1-\gamma_{t-1}} \geq 1$, ensuring $\mathcal{L}_\psi \geq 0$. In this paper, we choose to use a uniformly weighted loss and leave the exploration of the weighted loss to future work.

3.4 Experiments

All training data consists of low-resolution 20×20 images, setting $\sigma_{\mathbf{x}} = 0.92$ in units of low-resolution pixels, for consistency with common experimental conditions with a 60x magnification objective lens and numerical aperture (NA) of 1.4. We multiply ω_k by a constant $i_0 = 200$ for experiments for consistency with typical fluorophore emission rates. All KDEs have dimension 80×80 , are scaled between $[0, 1]$, and are generated using $\sigma_{\mathbf{y}} = 3.0$ pixels in the upsampled image. For a typical CMOS camera, this results in KDE pixels with lateral dimension of $\approx 27\text{nm}$. Initial coordinates θ were drawn uniformly over a two-dimensional disc with a radius of 7 low-resolution pixels.

Localization RMSE. In order to verify the initial predictions made by the augmentation model ϕ , we simulated a dataset $(\mathbf{x}_i, \mathbf{y}_{0,i}, \hat{\mathbf{y}}_i)_{i=1}^N$ with $N = 1000$. Objects in the KDE $\hat{\mathbf{y}}_i$ are detected using the Laplacian of Gaussian (LoG) detection algorithm [Kong2013], which permits more direct comparison of model predictions to the Cramer-Rao lower bound on localization error, compared to other image similarity measures. Localization is carried out from scale-space maxima directly in LoG, as opposed to fitting a model function to KDEs. A particular LoG localization in the KDE is paired to the nearest ground truth localization and is unpaired if a localization is not within 5 KDE pixels of any ground truth localization. In addition to localization error, we measure a precision $P = \text{TP}/(\text{TP} + \text{FP}) = 1.0$ and recall

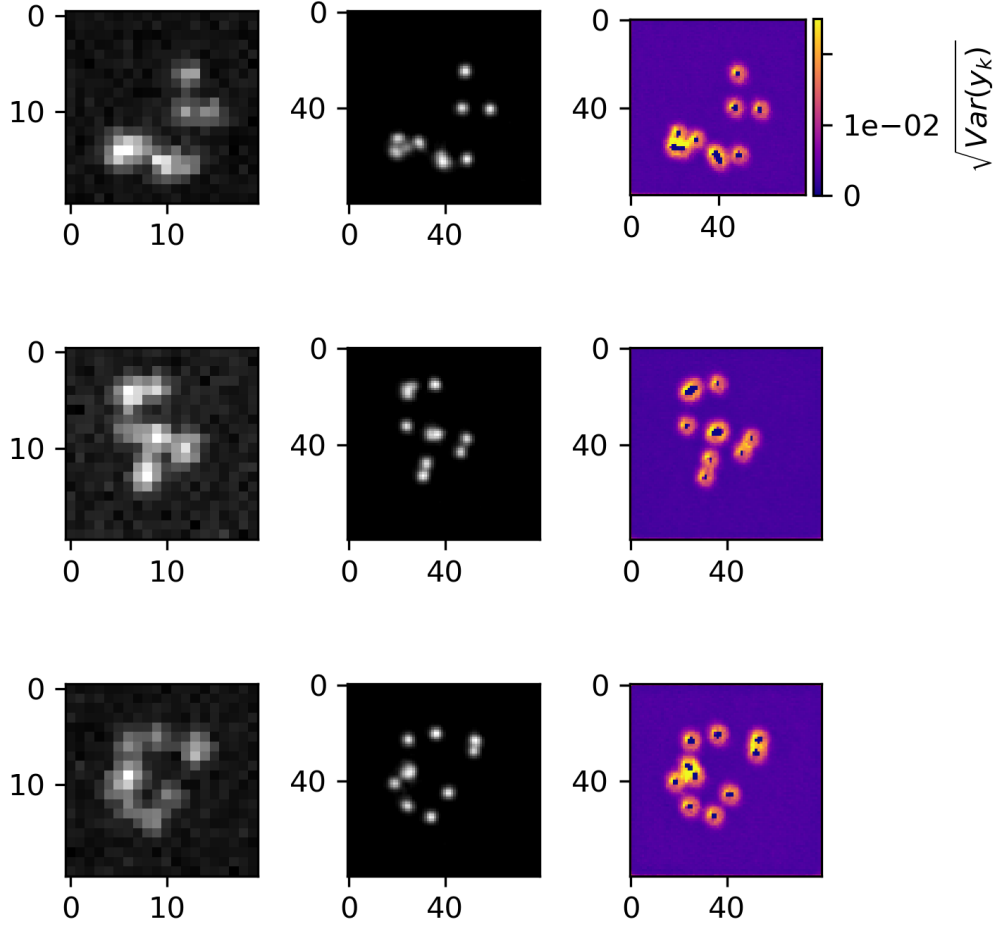


Figure 3.2. Non cherry-picked estimation of marginal variances. A low-resolution image \mathbf{x} (left column) is transformed by ϕ to produce a KDE estimate $\hat{\mathbf{y}}$ (middle column) and a DDPM ψ computes a map of marginal variances (right column)

$R = TP/(TP + FN) = 0.85$, where TP denotes true positive localizations, FP denotes false positive localizations, and FN denotes false negative localizations.

Variational Diffusion. We set $T = 100$ for all experiments and treat forward process variances β_t as hyperparameters, with a linear schedule from $\beta_0 = 10^{-4}$ to $\beta_T = 10^{-2}$. These constants were chosen to be small relative to ground truth KDEs, which are scaled to $[-1, 1]$, ensuring that forward process distribution $\mathbf{y}_T \sim q(\mathbf{y}_T|\mathbf{y}_0)$ approximately matches the reverse process $\mathbf{y}_T \sim \mathcal{N}(0, I)$ at $t = T$. Example KD estimates from low-resolution images and the marginal variances obtained from sampling $N = 100$ samples from $p_\psi(\mathbf{y}_0|\mathbf{x})$ are shown in (Figure 4).

3.5 Conclusion

We proposed a variational diffusion model for uncertainty-aware localization microscopy. Our approach builds on recent advancements in conditional diffusion models, to model the posterior distribution on high-resolution KD estimates from low-resolution inputs. This tractable posterior distribution is constructed by first augmenting low resolution inputs to a KD estimate using the DeepSTORM architecture with minor modifications [Nehme2020]. Conditioning a diffusion model on this initial estimate permits sampling with relatively fewer samples than most existing diffusion models in similar applications, thereby making computation of marginal variances possible. Our approach made three core contributions: (i) we derived a relationship between the posterior on kernel density estimates with the posterior on molecular locations, and (ii) we demonstrated that a diffusion model can model a distribution on KDEs with qualitatively similar marginal variances expected from predictions made using MCMC. By using a recently discovered relationship of the variational lower bound to a traditional noise-estimation objective, we can confidently approximate the true posterior.

3.6 Broader Impact

The development of a method for uncertainty estimation in super-resolution imaging, as proposed here, holds implications beyond its immediate application in SMLM. By leveraging

diffusion models for uncertainty estimation, this approach not only enhances the reliability of super-resolution image reconstructions but also extends its utility to a diverse array of domains. The incorporation of a guided diffusion process facilitates efficient reconstruction while maintaining interpretation of the underlying uncertainty. Importantly, the principles underlying this method resonate across various fields, suggesting its potential applicability in domains beyond microscopy. For instance, the extension of similar techniques to general image processing tasks highlights the potential to address uncertainty in a wide range of applications in bioimaging or medical imaging. Moreover, the utilization of diffusion models for uncertainty estimation aligns with a broader trend in leveraging probabilistic frameworks for enhancing deep learning applications, with implications extending to fields such as natural language processing, computer vision, and autonomous systems. By bridging these interdisciplinary boundaries, this method not only addresses a critical need in localization microscopy but also contributes to the advancement of uncertainty-aware deep learning methodologies.

3.7 Appendix

3.7.1 Sampling

Sampling from the reverse process $p_\psi(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{x})$ is achieved by estimation of the noise ϵ_ψ from \mathbf{y}_t by the denoising model ψ , and therefore estimation of \mathbf{y}_0

$$\hat{\mathbf{y}}_0 = \frac{1}{\sqrt{\gamma_t}}(\mathbf{y}_t - \sqrt{1 - \gamma_t}\epsilon_\psi) \quad (3.7)$$

followed by sampling from the forward process $\mathbf{y}_{t-1} \sim q(\mathbf{y}_{t-1}|\hat{\mathbf{y}}_0) = \mathcal{N}(\sqrt{\gamma_{t-1}}, (1 - \gamma_{t-1})I)$.

3.7.2 Derivation of the variational bound

We now derive the so-called diffusion loss \mathcal{L}_ψ , written in (8) in the main text. Similar derivations can be found in [Kingma2021, Ribeiro2024], and we include it here only for completeness

$$\begin{aligned}
-\log p(\mathbf{y}_0) &\leq -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \frac{p(\mathbf{y}_{0:T})}{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \\
&= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \log \frac{p(\mathbf{y}_T)p(\mathbf{y}_0|\mathbf{y}_1) \prod_{t=2}^T p(\mathbf{y}_{t-1}|\mathbf{y}_t)}{q(\mathbf{y}_T|\mathbf{y}_0) \prod_{t=2}^T q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)} \\
&= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} \left[p(\mathbf{y}_0|\mathbf{y}_1) + \log \frac{p(\mathbf{y}_T)}{q(\mathbf{y}_T|\mathbf{y}_0)} + \sum_{t=2}^T \log \frac{p(\mathbf{y}_{t-1}|\mathbf{y}_t)}{q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)} \right] \\
&= -\mathbb{E}_{q(\mathbf{y}_{1:T}|\mathbf{y}_0)} [p(\mathbf{y}_0|\mathbf{y}_1)] + D_{KL}(q(\mathbf{y}_T|\mathbf{y}_0)||p(\mathbf{y}_T)) \\
&\quad + \sum_{t=2}^T \mathbb{E}_{q(\mathbf{y}_t|\mathbf{y}_0)} D_{KL}(q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)||p(\mathbf{y}_{t-1}|\mathbf{y}_t))
\end{aligned}$$

As before, we omit conditioning on \mathbf{x} to simplify the notation. The first term is typically ignored, as it does not contribute meaningfully to the loss [Ribeiro2024]. Furthermore, the second term is approximately zero by construction. Therefore we are left with the last term, called the diffusion loss \mathcal{L}_ψ . The KL-divergence of q and p is between two Gaussians with identical variances $\sigma^2 = \frac{(1-\gamma_{t-1})(1-\alpha_t)}{1-\gamma_t}$, and expectations

$$\mu = \frac{\sqrt{\gamma_{t-1}}(1-\alpha_t)}{1-\gamma_t} \mathbf{y}_0 + \frac{\sqrt{\alpha_t}(1-\gamma_{t-1})}{1-\gamma_t} \mathbf{y}_t \quad \mu_\psi = \frac{\sqrt{\gamma_{t-1}}(1-\alpha_t)}{1-\gamma_t} \hat{\mathbf{y}}_0 + \frac{\sqrt{\alpha_t}(1-\gamma_{t-1})}{1-\gamma_t} \mathbf{y}_t$$

for a fixed noise schedule [Saharia2021]. Therefore, we have

$$\begin{aligned}
D_{KL}(q(\mathbf{y}_{t-1}|\mathbf{y}_t, \mathbf{y}_0)||p(\mathbf{y}_{t-1}|\mathbf{y}_t)) &= \frac{1}{2\sigma^2} \|\mu - \mu_\psi\|_2^2 \\
&= \frac{1}{2} \frac{\gamma_{t-1}(1-\alpha_t)}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\
&= \frac{1}{2} \frac{\gamma_{t-1}((1-\gamma_t) - \alpha_t(1-\gamma_{t-1}))}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\
&= \frac{1}{2} \frac{\gamma_{t-1}((1-\gamma_t) - \frac{\gamma_t}{\gamma_{t-1}}(1-\gamma_{t-1}))}{(1-\gamma_{t-1})(1-\gamma_t)} \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\
&= \frac{1}{2} \left(\frac{\gamma_{t-1}}{1-\gamma_{t-1}} - \frac{\gamma_t}{1-\gamma_t} \right) \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 \\
&= \frac{1}{2} (\text{SNR}_{t-1} - \text{SNR}_t) \|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2
\end{aligned}$$

Reparameterizing the loss in terms of the noise, using $\|\mathbf{y}_0 - \hat{\mathbf{y}}_0\|_2^2 = \frac{1-\gamma_t}{\gamma_t} \|\epsilon_0 - \epsilon_\psi\|_2^2$ [Ribeiro2024], we arrive at

$$\mathcal{L}_\psi = \frac{1}{2} \sum_{t=2}^T \mathbb{E}_{q(\mathbf{y}_t|\mathbf{y}_0)} \left(\frac{\text{SNR}_{t-1}}{\text{SNR}_t} - 1 \right) \|\epsilon - \epsilon_\psi\|_2^2$$

Using a Monte Carlo estimate of \mathcal{L}_ψ [Kingma2023] which optimizes random terms of the summation to avoid calculating all terms simultaneously, we arrive at the objective written in the main text (8)

$$\mathcal{L}_\psi = \mathbb{E}_{\epsilon \sim \mathcal{N}(0, I), t \sim U(1, T)} \left[\left(\frac{\text{SNR}_{t-1}}{\text{SNR}_t} - 1 \right) \|\epsilon - \epsilon_\psi\|_2^2 \right]$$

3.7.3 Metropolis-Hastings MCMC

To obtain numerical estimates of $p(\theta|\mathbf{x}) \propto p(\mathbf{x}|\theta)p(\theta)$ and therefore $p(\mathbf{y}|\mathbf{x})$, for the isolated fluorescent molecule as shown in (Figure 2), we used Metropolis-Hastings Markov Chain Monte Carlo (MCMC) to estimate the posterior on coordinates. Under the Poisson approximation in (1), the model negative log-likelihood is

$$\ell(\mathbf{x}|\theta) = -\log \prod_k \frac{e^{-(\omega'_k)} (\omega'_k)^{n_k}}{n_k!} = \sum_k \log n_k! + \omega'_k - n_k \log (\omega'_k) \quad (3.8)$$

where n_k is the observed number events at a pixel. MCMC is asymptotically exact, which is not guaranteed by variational methods which may rely on a Laplace approximation around the MLE. We choose a uniform prior $p(\theta)$, and Metropolis-Hastings is run for 10^4 iterations, the first 10^3 iterations are discarded as burn-in. A proposal $\theta' = \theta + \Delta\theta$ was generated with $\Delta\theta \sim \mathcal{N}(0, \sigma^2 I)$ where $\sigma^2 = 0.05$. The acceptance probability is

$$\alpha = e^{\beta(\ell(\theta) - \ell(\theta'))}$$

We choose $\beta = 0.2$ to achieve a target acceptance rate of 0.5.

3.7.4 Cramer-Rao Lower Bound

Reliable inference of θ from \mathbf{x} in general requires performance metrics for model selection. We use the Fisher information as an information theoretic criteria to assess the quality of the data augmentation model ϕ tested here, with respect to the root mean squared error (RMSE) of our predictions of θ [Chao2016]. The Poisson log-likelihood $\ell(\mathbf{x}|\theta)$ is also convenient for computing the Fisher information matrix [Smith2010] and thus the Cramer-Rao lower bound, which bounds the variance of a statistical estimator of θ , from below i.e., $\text{var}(\hat{\theta}) \geq I^{-1}(\theta)$. The Fisher information is straightforward to compute under the Poisson log-likelihood in (1). In general, the Fisher information is given by the expression

$$I_{ij}(\theta) = \mathbb{E}_{\theta} \left(\frac{\partial \ell}{\partial \theta_i} \frac{\partial \ell}{\partial \theta_j} \right) \quad (3.9)$$

For an arbitrary parameter, we find that, for a Poisson log-likelihood ℓ

$$\begin{aligned} \frac{\partial \ell}{\partial \theta_i} &= \frac{\partial}{\partial \theta_i} \sum_k \log n_k! + \omega'_k - n_k \log(\omega'_k) \\ &= \sum_k \frac{\partial \omega'_k}{\partial \theta_i} \left(\frac{\omega'_k - n_k}{\omega'_k} \right) \end{aligned}$$

Using this result, we can compute the Fisher information matrix $I(\theta)$

$$I_{ij}(\theta) = \mathbb{E}_{\theta} \left(\sum_k \frac{\partial \omega'_k}{\partial \theta_i} \frac{\partial \omega'_k}{\partial \theta_j} \left(\frac{\omega'_k - n_k}{\omega'_k} \right)^2 \right) = \sum_k \frac{1}{\omega'_k} \frac{\partial \omega'_k}{\partial \theta_i} \frac{\partial \omega'_k}{\partial \theta_j}$$

A fundamental lower bound on the variance in our estimates of θ then is found from its inverse: $\text{CRLB} = I^{-1}(\theta)$. This result is used to show in (Figure 5), that the data augmentation model ϕ efficiently estimates molecular coordinates under the experimental conditions tested here.

3.7.5 Neural Networks ψ, ϕ

DeepSTORM CNN ϕ . The DeepSTORM CNN, for 3D localization, can be viewed as a deep kernel density estimator, reconstructing kernel density estimates \mathbf{y} from low-resolution

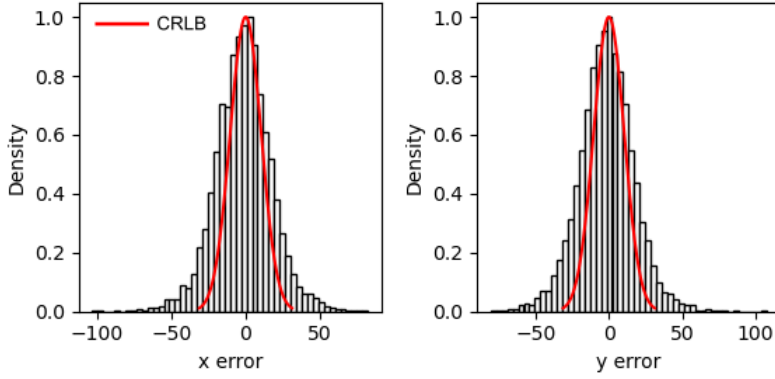


Figure 3.3. Localization errors of the trained model ϕ . The Cramer-Rao lower bound is shown in red, computing by taking the diagonal elements of $I^{-1}(\theta)$.

inputs \mathbf{x} . We utilize a simplified form of the original architecture [Nehme2020] for 2D localization, which we denote ϕ in this paper, which consists of three main modules: a multi-scale context aggregation module, an upsampling module, and a prediction module. For context aggregation, the architecture utilizes dilated convolutions to increase the receptive field of each layer. The upsampling module is then composed of two consecutive 2x resize-convolutions, computed by nearest-neighbor interpolation, to increase the lateral resolution by a factor of 4. Additional details regarding this architecture can be found in the original paper Nehme2020. The terminal prediction module contains three additional convolutional blocks for refinement of the upsampled image, followed by an element-wise HardTanh. The architecture is trained using the objective $\mathcal{L}_\phi = \frac{1}{N} \sum_{n=1}^N (\mathbf{y}_{0,n} - \hat{\mathbf{y}}_n)^2$.

DDPM ψ . To represent the reverse process, we used a DDPM architecture originally proposed in [Saharia2021]. We chose the U-Net backbone to have channel multipliers [1, 2, 4, 8, 8] in the downsampling and upsampling paths of the architecture. In this architecture, parameters are shared across time, which is specified to the network using the Transformer sinusoidal position embedding, and uses self-attention at the 16×16 feature map resolution. To condition the model on the input $\hat{\mathbf{y}}$, we concatenate the $\hat{\mathbf{y}}$ estimated by DeepSTORM along the channel dimension, which are scaled to $[0, 1]$, with $\mathbf{y}_T \sim \mathcal{N}(0, I)$.

Others have experimented with more sophisticated methods of conditioning, but found that the simple concatenation yielded similar generation quality [**Saharia2021**].

4. Enhanced SMLM with single photon counting cameras

Conventional single molecule localization microscopy infers the locations of putatively isolated fluorescent emitters to produce super-resolved images. The number of active fluorescent emitters is therefore a critical piece of information during localization; however, this quantity is generally underdetermined in widefield fluorescent microscopy. Nevertheless, many fluorophores used in bioimaging are intrinsically single photon sources which exhibit fluorescence antibunching and can emit only one photon following a very short excitation pulse. This general property may provide the necessary physical information for constrained localization in non-sparse scenes by yielding estimates of the number active fluorescent emitters. Here, we leverage recent advancements in single photon avalanche diode (SPAD) arrays to accurately count the number active emitters in a fluorescent sample, while concurrently performing intensity-based multi-emitter localization. We find that SPAD cameras, with their high temporal resolution and single photon sensitivity are capable of precise single molecule localization microscopy, while maintaining the relative simplicity of widefield imaging.

4.0.1 Introduction

Far-field optical microscopy is fundamentally limited by diffraction, with the maximum attainable resolution being limited to approximately half the wavelength of light. Several schemes to beat the diffraction limit have been developed in recent years. Many of these schemes utilize the concept of precise localization of isolated fluorescent emitters which blink over a time series of frames [Rust2006, Betzig2006]. An inherent problem with such methods is the requirement that fluorescent emitters be isolated, slowing down the acquisition of super-resolved images. To address this, we leverage the fact that many fluorophores are intrinsically single photon sources and exhibit fluorescence antibunching. This property can constrain the number of active fluorescent emitters in a region of interest (ROI) and can potentially enable localization in non-sparse scenes [Ta2010, Israel2017].

Molecular counting with fluorescence antibunching has a fairly simple motivation: coincidence of photons at multiple detector elements during high speed imaging provides evidence for the number of single photon sources present in the imaged region. Combining the ideas of

conventional super-resolution approaches with photon statistics may prove to be a powerful set of methods for bioimaging. SPAD cameras achieve orders of magnitude higher temporal resolutions than standard CMOS cameras, single photon sensitivity, and dark count rates less than 25cps. Furthermore, the reduced readout noise and large fill-factor of recently commercialized SPAD arrays suggests their use for single molecule localization with reduced localization uncertainty. Localization uncertainty, typically the root mean square error (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 [Chao2016]. Managing the increase in localization uncertainty at high labeling density remains a major bottleneck to localization microscopy. For example, static uncertainty due to molecular crowding can be partially ameliorated by using pairwise or higher-order temporal correlations within a pixel neighborhood [Dertinger2009]. However, the number of fluorescent active emitters in a region of interest remains critical prerequisite information in single molecule localization.

In this study, we present a method for widefield single photon counting in order to rigorously count fluorophores in the sample and subsequently constrain single molecule localization. We investigate the theoretical properties of the zero-lag second-order coherence function $g^{(2)}(0)$ for widefield photon counting and its spatial properties. Using Bayesian analysis, we derive a posterior distribution on the number of active fluorescent emitters in a region of interest. We then combined this with single molecule localization algorithms and demonstrate resolution of multiple emitters using a multi-emitter fitting algorithm and report localization errors with respect to the Cramer-Rao bound.

4.0.2 Basic Scheme

We consider a simplified description of widefield photon counting for a single photon source in the object plane labeled by a continuous-valued coordinate $\theta = (x_0, y_0)$. The spatial profile O of the field in image space is presumed to have a Gaussian shape [Zhang2007, Richards1959, Gibson1989].

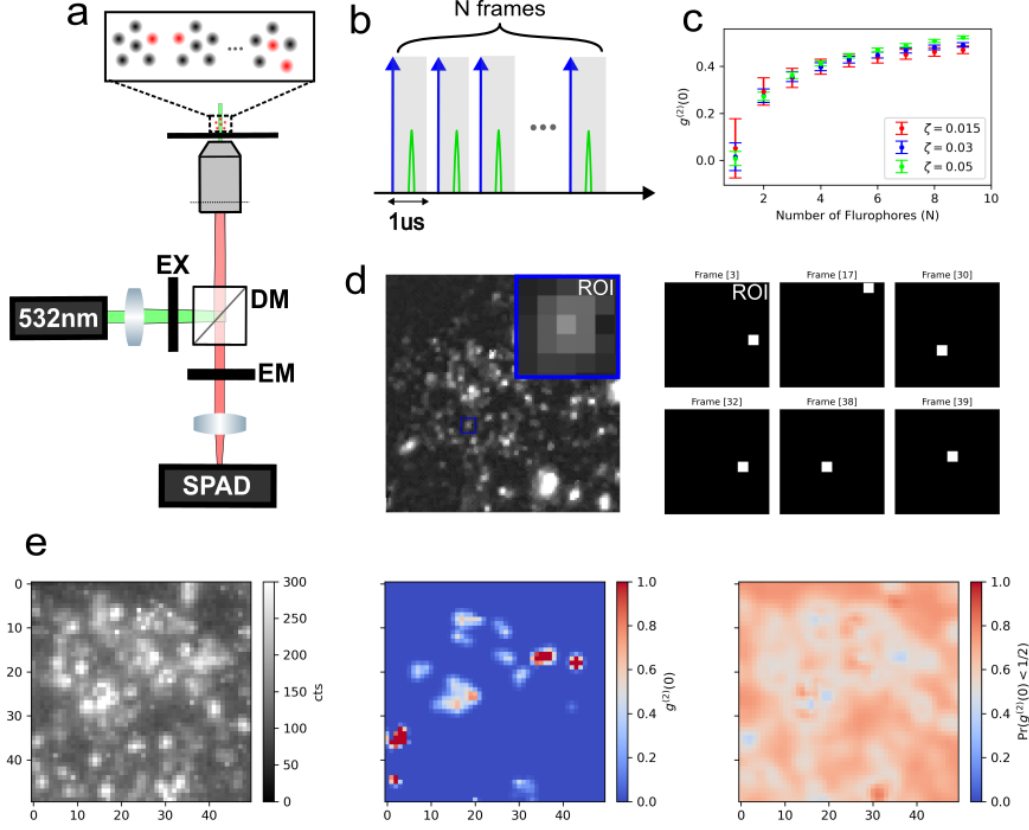


Figure 4.1. Single photon counting with a SPAD array (a) Conventional widefield microscopy with integrated SPAD array (b) Single photon imaging scheme using 1 μ s exposures containing a picosecond laser pulse (c) Sum of photon counts over a 5x5 region of interest (ROI), taken with $N_{\text{frames}} = 5 \times 10^5$

$$O(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}} \quad (4.1)$$

Therefore the field operator in object space is $\hat{E} \propto \hat{a}$ and in image space $\hat{E} \propto O(x, y)\hat{a}$. Since our SPAD detectors at the image plane must be discrete, the total field at a detector element k centered in image space at $s_k = (u_k, v_k)$ is then given by integrating over pixels of width δ . Moreover, the Gaussian O is presumed to be isotropic and therefore we have $\hat{E}(s_k) \propto \Gamma_x(u_k, x_0)\Gamma_y(v_k, y_0)$. For example,

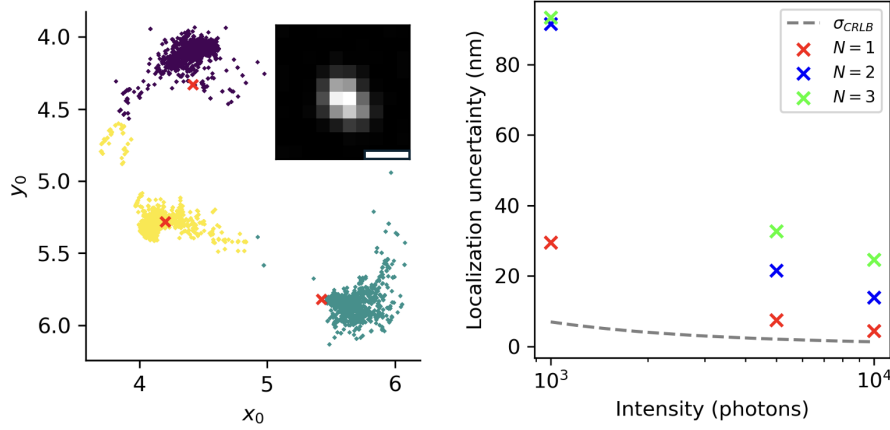


Figure 4.2. Single and multi-emitter localization error on sums of photon counts. (left) Localization uncertainty for simulated data for different values of N , plotted with respect to the Cramer-Rao lower bound, shown in dashed gray. (right) Multi-emitter localization by MCMC sampling for $N = 3$, colors indicate a cluster of samples i.e., a single localization. All data was generated with a background rate $\langle \mathbf{n}_{\text{background}} \rangle = \lambda N_{\text{frames}}/d^2$ per pixel. Scalebar 360nm

$$\Gamma_x(u_k, x_0) = \frac{1}{\sqrt{2}} \left(\text{erf} \left(\frac{u_k + \frac{1}{2} - x_0}{\sqrt{2}\sigma} \right) - \text{erf} \left(\frac{u_k - \frac{1}{2} - x_0}{\sqrt{2}\sigma} \right) \right)$$

We now consider the case of pulsed excitation where the interval between pulses much longer than the fluorescence lifetime. Upon excitation of an isolated fluorescent emitter, a photon is detected at a particular detector element k with probability $\zeta_k \propto \langle \hat{E}^\dagger(s_k) \hat{E}(s_k) \rangle = \frac{1}{2} \Gamma_x^2 \Gamma_y^2 \text{Tr}(\rho a^\dagger a)$ where ρ is the density matrix for a two-level system. Similarly, the probability of detection in a region of interest collecting all photons emitted is $\zeta \propto \text{Tr}(\rho a^\dagger a)$. Here, we are primarily concerned with the latter quantity, and its application in counting fluorescent emitters.

By temporarily ignoring the spatial profile described by (1), we derive a likelihood on the number of fluorophores in a small ROI with a lateral dimension $d = 5$ pixels. For N fluorophores emitting photons which can be detected within a ROI of the SPAD array, the number of signal photons measured n_{signal} following a single excitation pulse will have Bi-

nomial statistics $n_{\text{signal}} \sim \text{Binom}(N, \zeta)$. Photon pile-up at a single detector element can be safely neglected in this model due to its relatively low likelihood. We then model the background signal within the region of interest as a coherent state, which must follow Poissonian statistics $n_{\text{background}} \sim \text{Poisson}(\lambda)$ for an expected number λ of background counts in the ROI per frame. The total number of counts $n = n_{\text{signal}} + n_{\text{background}}$ detected in the region of interest following a single pulse is then distributed by the likelihood

$$p(n = n' \mid N, \zeta) = \sum_{i=0}^{\infty} \binom{N}{i} \zeta^i (1 - \zeta)^{N-i} \frac{\lambda^{n'-i}}{(n' - i)!} e^{-\lambda} \quad (4.2)$$

The expression in (2) represents a convolution of Poisson and Binomial probability mass functions. This result is the primary means of inference of the number of active emitters N in a ROI.

In order to begin to perform localization in non-sparse ROIs, we write a posterior distribution on the Binomial parameters used in the likelihood (2) using Bayes rule

$$p(N, \zeta \mid x) \propto p(x \mid N, \zeta) p(\zeta) \quad (4.3)$$

We use a Gaussian prior on ζ i.e., $p(\zeta) = \mathcal{N}(\mu_{\zeta}, \sigma_{\zeta})$ with $\mu_{\zeta} = 0.01$ and $\sigma_{\zeta} = 0.005$. Prior uncertainty in the value of ζ stems from fluorophores with potentially heterogeneous photophysical properties as well as varying laser power throughout the excited region. This posterior can be integrated over ζ to produce a posterior distribution on the fluorophore number N i.e., $p(N = N' \mid n) \propto \int_0^1 \prod_j p(n_j \mid N', \zeta) p(\zeta) d\zeta$ which can be estimated using Monte Carlo methods. The final posterior is then estimated by minibatching the data into batches of 10^3 frames and averaging the posterior $p(N \mid n)$ over minibatches. The fluorophore number N within each ROI is then estimated by the maximum a posteriori (MAP) estimate N^* given by this distribution.

For localization, we notice that (2) is well approximated by a Poisson distribution for a large frame number, making the localization procedure similar to conventional intensity-based methods [Smith2010]. Denoting the fluorophore coordinates by θ and vector of total counts in the region of interest \mathbf{n} , we have the following log-likelihood

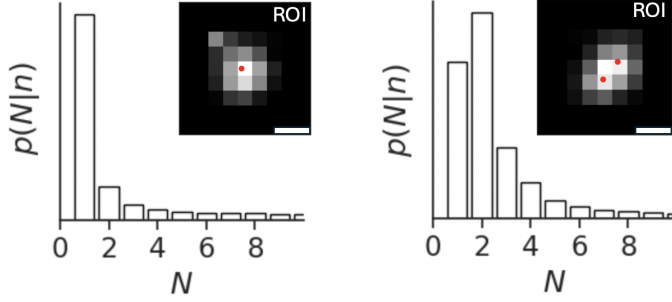


Figure 4.3. Posteriors on the number of fluorescent emitters N and localization for $N^* = 1$ (left) and $N^* = 2$ (right) quantum dots. Scalebars 360nm

$$\ell(\mathbf{n}|\theta) = -\log \prod_k \frac{e^{-(\mu_k)} (\mu_k)^{\mathbf{n}_k}}{\mathbf{n}_k!} \quad (4.4)$$

$$= \sum_k \log \mathbf{n}_k! + \mu_k - \mathbf{n}_k \log (\mu_k) \quad (4.5)$$

where, in the multi-emitter regime the expected photon count at a pixel is $\mu_k = \langle \mathbf{n}_k \rangle = \sum_{m=1}^{N^*} \mu_{k,m}$ given $\mu_{k,m} = \zeta N_{\text{frames}} \Gamma_x(u_k, x_{0,m}) \Gamma_y(v_k, y_{0,m}) + \lambda N_{\text{frames}} / d^2$. In the multi-emitter regime, optimization of (4) by sampling is a suitable choice (see Results).

4.0.3 Results

Quantum dots coated on a glass coverslip were excited using a picosecond 532nm pulsed laser triggered at 500kHz. Emission light was collected using an oil-immersion 100 \times objective with numerical aperture (NA) 1.4 (Nikon). The emission signal was then filtered to exclude the laser line (Semrock) and projected onto the SPAD512 sensor (Pi Imaging Technologies) using a tube lens. A simplified diagram of the complete system is depicted in (Figure 1a). Each acquisition consists of $N = 5 \times 10^5$ frames (500ms), synchronized with each laser pulse, using a 1 μ s exposure per frame (Figure 1b,d). To confirm the presence of single photon sources in the sample, we investigated properties of the zero-lag second order coherence function $g^{(2)}(0)$. The following empirical estimate of $g^{(2)}(0)$ is used [Israel2017]

$$g^{(2)}(0) = \frac{G^{(2)}(0) - B}{\langle G^{(2)}(m \neq 0) \rangle - B} \quad (4.6)$$

where $B = N_{\text{frames}}\lambda\zeta$ is the expected number of background-signal coincidences in the region of interest. The quantity $G^{(2)}(m)$ represents the number of signal-signal coincidences in the region of interest at a lag time m . The quantity $\langle G^{(2)}(m \neq 0) \rangle$ is the average number of coincidences in pairs of frames at nonzero lag $m \in [1, 100]$, in units of frames. As expected, simulation from the likelihood (2) shows saturation of $g^{(2)}(0)$ with increasing values of N (Figure 1c). Moreover, bright clusters of quantum dots exhibit elevated $g^{(2)}(0)$ values as can be seen in maps of the $g^{(2)}(0)$ computed using sliding window over the array (Figure 1e).

For localization by optimization of (4), we use Goodman and Weare's Markov Chain Monte Carlo (MCMC) algorithm [Goodman2010] to sample from the posterior on fluorophore locations. In all simulations we assume a uniform prior on coordinates over the ROI and ζ is known and identical over fluorophores. Fluorophore locations can then be estimated from the posterior samples by K-means clustering of the (x, y) coordinates of the first particle and identification of cluster centers (Figure 2a). To validate our estimator, we compare its RMSE to the single emitter Cramer-Rao lower bound, which bounds the variance of a statistical estimator of θ , from below (Chao 2016). For an isolated emitter, the Poisson log-likelihood (5) is convenient for computing the Fisher information matrix for θ and thus the Fisher information is [Smith2010]

$$I_{ij}(\theta) = \sum_k \frac{1}{\mu_k} \frac{\partial \mu_k}{\partial \theta_i} \frac{\partial \mu_k}{\partial \theta_j} \quad (4.7)$$

The Cramer-Rao bound is then found by $\text{var}(\theta) \geq I^{-1}(\theta)$. We find that for thousands of expected signal photon counts, localization uncertainty lies in an acceptable range for localization microscopy (Figure 2b). Example posteriors and multi-emitter fitting on experimental quantum dot data are found in (Figure 3). The ζ value is treated as unknown but homogeneous across fluorophores in the ROI. *A major criticism is likely that we have not measured ζ precisely and that we can only provide self-consistency for its simultaneous value with N .*

4.0.4 Discussion

Many fluorescent emitters exhibit random variations of brightness known as blinking. Blinking increases the observed photon-number fluctuations and could be expected to affect the value of $g^{(2)}(0)$ or the posterior on the number of active fluorescent emitters. However, the signal photon number per frame will follow Binomial statistics even in the presence of blinking, the only consequence of which is an effective reduction of the detection probability ζ . If the effect of censoring photons by blinking and lowering the quantum yield can be accounted for, the technique used here may be compatible with common super-resolution techniques such as stochastic optical reconstruction microscopy (STORM).

The acquisition times necessary to obtain sufficient photon counts for computing the necessary statistics can potentially be very short. Most fluorophores have relaxation times in the nanosecond range and thus photons can be collected at a rate of at tens of millions of excitation pulses per second. These rates are currently difficult to obtain, however, due to limitations in detector throughput. The SPAD camera used in this study has a minimum exposure time in the microsecond range. Furthermore, the data volume can quickly become intractable due to the need for several thousands of frames for a millisecond-scale exposure time. This is currently a complication for techniques like STORM and advancements in the automation for data acquisitions are necessary. The speed of MCMC based localization remains a limitation for post-processing, and optimization of the processing time for localization is left for future work.

In conclusion, we propose a single molecule imaging technique that allows for simultaneous counting of localization of fluorescent molecules by modeling the quantum properties of fluorescence emission. The technique does not require a nonclassical light source and is designed to supplement standard single molecule localization microscopy techniques. The proposed method can be implemented with a standard widefield fluorescence microscope.