9-1-2017

Statistical Models for Single Molecule Localization Microscopy

Ahmed Elmokadem
University of Connecticut - Storrs, elmokadem82@gmail.com

Statistical Models for Single Molecule Localization Microscopy

Ahmed Shawky Elmokadem, PhD

University of Connecticut, 2017

Single-molecule localization microscopy (SMLM) has revolutionized the field of cell biology. It allowed scientists to break the Abbe diffraction limit for fluorescence microscopy and got it closer to the electron microscopy resolution but still it faced some serious challenges. Two of the most important of these are the sample drift and the measurement noise problems that result in lower resolution images. Both of these problems are generally unavoidable where the sample drift is a natural mechanical phenomenon that occurs during the long time of image acquisition required for SMLM (Geisler et al. 2012) while the measurement noise, which arises from combining localization uncertainty and counting error, is related to the number of photons collected from the fluorophore and affects the precision in locating the centroids of single molecules (Thompson, Larson, and Webb 2002).

Previous work has tried to devise methods to deal with the sample drift problem but unfortunately, these methods either add too much complexity to the experimental setup or are just inefficient in correctly estimating the drift at the single frame level (Wang et al. 2014). As for measurement noise, all current regular image rendering algorithms treat every detection of the fluorophore as a separate event and hence, the localization uncertainty of every detection of the same molecule would give offset coordinates from the other detections leading to a distorted final image.

In this thesis, I demonstrate two novel approaches based on statistical concepts to address each of these two problems. The algorithm for solving the sample drift problem is based on

Ahmed Shawky Elmokadem, PhD

University of Connecticut, 2017

Bayesian inference and it showed efficiency in estimating drift at the single-frame level and proved superior and more straightforward than the available methods. The algorithm for addressing the measurement noise problem is based on Gibbs sampling and not only did it enhance resolution, but it also offers for the first time a means to quantify resolution uncertainty as well as uncertainty in cluster size measurement for clustering proteins. Therefore, this work offers a significant advancement in the field of SMLM and more generally, cell biology.

Statistical Models for Single Molecule Localization Microscopy

Ahmed Shawky Elmokadem

B.S., Misr International University, 2005

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2017

APPROVAL PAGE

Doctor of Philosophy Dissertation

Statistical Models for Single Molecule Localization Microscopy

Presented by

Ahmed Shawky Elmokadem, B.S.

Major Advisor

_____

Ji Yu

Co-Major Advisor

_____

Ion Moraru

Associate Advisor

_____

Bruce Mayer

Associate Advisor

_____

Yi Wu

Associate Advisor

_____

Ann Cowan

Associate Advisor

_____

John Carson

University of Connecticut

2017

Acknowledgements

First I would like to thank my major advisors, Ji Yu and Ion Moraru. Ji, I still remember your first words to me when I approached you for a rotation in your lab 4 years ago. Those words were: I am not looking for students right now; but for some reason you still took me in after the rotation was over and I cannot imagine what I would have done without you. I was able to do the kind of unconventional work that I have done for this thesis only because your vast knowledge about so many fields made it possible. You taught me a new thing every day and were so patient with all of my stupid mistakes. Besides the scientific guidance I got from you, you also gave me tremendous personal support when I needed it the most during the very bad phase I had a couple of years ago and for that, I am forever in your debt. Ion, the thing about you is that no matter how brief our meetings were, you always managed to be on top of things and to give me exactly what I need to keep going. Thank you both for putting me on the path I am in right now.

I would like to thank my amazing committee members: Bruce Mayer, Yi Wu, Ann Cowan and John Carson. Bruce and Yi, our lab's extensive collaboration with you brought your unique insights into my project and my progress. I was privileged to have worked in both of your labs briefly and to learn from you directly about concepts and experimental procedures that I was learning for the first time. Ann and John, thank you for all of our fruitful discussions and for being always there for me and other students. Besides being great scientists, you also have this gentle personal quality that you know how to push students to fulfill their potential without making them too stressed about it.

To my past and current lab members and colleagues: Cliff, Sulagna, Qing, Eunji, Massoud and Abhi. We have been through a lot together guys and I will always remember the fun times we had working together and our discussions that occasionally deviated from cell biology to philosophy and the purpose of the universe.

To Michael Blinov. You were the first to introduce me to the amazing world of CCAM and to mathematical modeling. I remember how I messed up the first rate equation you asked me to write down but you didn't give up on me and walked me patiently through the process of how to think quantitatively about everything.

To my incredible mentor, Dr. Karim Kamal. You taught me never to be satisfied with where I am and to always work on developing myself. It was you who opened up the world of real science in front of me and were always backing my decisions even if others doubted me. Thank you Dr. Karim for having such an inspiring spirit that not even illness can extinguish.

My deepest appreciation and gratitude goes to Dr. Catherine Wu and Dr. George Wu. You were the first to introduce me to the Health Center and to the amazing research being carried out here and you were also integral to my transition from Egypt to living in Connecticut. Your support and help inspired me and shaped me in the way I am right now.

To my parents who made every effort for me to be here. To my mother who inflamed the passion for learning in my heart and who set an example for me to follow in pursuing an academic career. To my father who has been my best friend and backbone in this journey that is life. Although it wasn't always easy but you guys made it your task to provide me with the best education and life you could possibly make and I hope I made you proud.

A special thanks has to go to Kimberly Dodge-Kafka. I still remember our phone interview before I got accepted into the program. You had a bad sore throat and your voice was all gone that we needed someone to translate but still it was a really fun interview and apparently you gave a good word on my behalf because here I am. You were also a huge support for me during the tough first year and one of the most important things I learned from you during the old Logic class was how to think like a real scientist. Thank you Kim for everything.

Finally, I would like to thank my beautiful wife and son who I dedicate this work to. Zina, you have been there for this whole journey and boy it was a difficult one but your faith and confidence were what kept me going. Thank you for bearing with me and for giving me a life I could only dream of. Adam, you are probably too young now to understand but I hope that this moment would make you someday proud of your dad.

Table of Contents

List of Figures

Abbreviations

NA – Numerical aperture
PSF – Point spread function
FWHM – Full width at half-maximum
NSOM – Near-field optical scanning microscopy
STED – Stimulation emission depletion
SSIM – Saturated structured illumination microscopy
SMLM – Single-molecule localization microscopy
PALM – Photoactivated localization microscopy
STORM – Stochastic optical reconstruction microscopy
FPALM – Fluorescence photoactivation localization microscopy
PA-FP – Photoactivatable fluorescent protein
PA-GFP – Photoactivatable green fluorescent protein
EosFP – Eos fluorescent protein
UV – Ultraviolet
PAINT – Point accumulation for imaging in nanoscale topography
MAP – Maximum a posteriori
EM – Expectation maximization
MLE – Maximum likelihood estimation
BaSDI – Bayesian sample drift inference
HMM – Hidden Markov model
MCMC – Markov chain Monte Carlo
CCAM – Center for Cell Analysis and Modeling

Chapter 1

Introduction

1.1 Super-resolution Imaging

The field of cell biology made huge advancements following the discovery of fluorescence microscopy in the early twentieth century (Masters 2010). This method had the advantage over other conventional optical microscopy methods in that it provided a much higher contrast and visibility allowing for better studying of sub-cellular structures. Furthermore, with the introduction of antibody-conjugated fluorescent probes and genetically-engineered fluorescent protein constructs, molecule-specific labeling became possible (Huang, Bates, and Zhuang 2010; Rust, Bates, and Zhuang 2006; Heintzmann and Ficz 2013; Masters 2010). This was a huge step for cell biologists since it allowed them to decipher the internal machinations of cells in an unprecedented manner.

Nevertheless, fluorescence microscopy remained for a long time plagued by a resolution barrier that faces all optical microscopy techniques. This barrier is called the "Abbe diffraction limit" because it was popularized by the German physicist, Ernst Abbe, in 1873 when he coined his famous mathematical equation that described the effect of light diffraction on resolution (Abbe 1873):

$$d = \frac{\lambda}{2n\ sin\theta}$$

where d represents the minimum resolvable distance between two objects, i.e., resolution, λ is the wavelength of light used for imaging, n is the refractive index of the medium through which light is traveling at an angle θ and the term $n\ sin\theta$ is also referred to as the numerical aperture (NA) and is a property of the objective lens used. Since imaging a point object will give an intensity distribution known as the point spread function (PSF), the resolution d can also be

thought of as the FWHM (full width at half-maximum) of that PSF. Taking into consideration the fact that almost all optical microscopy techniques use visible light as the source of illumination with wavelengths that fall roughly in the range between 400 and 700 nm, and assuming a good NA, we can use the previous equation to calculate the maximum achievable lateral spatial resolution for optical microscopy methods, which will be about 200-300 nm.

This Abbe diffraction limit stood as a real challenge for cell biologists studying protein localizations and processes taking place at the level of ultra-structures with much smaller sizes. Among these many structures are single microtubules or actin filaments whose widths are about 25 and 6 nm, respectively (Fallis 2013), or nanoclusters formed by some membrane proteins with diameters of only a few tens of nanometers (Garcia-parajo et al. 2014).

In 1931, electron microscopy was developed where accelerated electrons were used as the source of illumination instead of light. This technique allowed for reaching much better resolutions than those attained by optical microscopy techniques since electrons have a much shorter wavelength than light and hence, can resolve much smaller structures (Giepmans 2008). Some of the more recent advancements in electron microscopy can even reach an astounding resolution of 50 pm, which is 4000 times lower than the Abbe diffraction limit for optical microscopy (Erni et al. 2009). Nevertheless, this couldn't sate the need for a fluorescence-based super-resolution microscopy method. This is because electron microscopy-based methods have lower specificity, higher purchase and maintenance costs and a much more tedious sample preparation process compared with fluorescence microscopy. In addition, the sample preparation and the high energy of the electron beam make it impossible to image live cells, which is something fluorescence microscopy can readily achieve (Huang, Bates, and Zhuang 2010).

The effort towards achieving a fluorescence microscopy method that could reach a sub-diffraction resolution finally bore fruit in 1986 when the first super-resolution fluorescence images were taken using a technique called near-field scanning optical microscopy (NSOM) (E. Betzig et al. 1986). This technique depends on placing a detector at a considerably shorter distance to the sample than the wavelength of the exciting light rendering the aperture size the only factor limiting resolution instead of diffraction. The resolution achieved by this technique was lower than 150 nm but unfortunately, NSOM suffered from many limitations; most notably it is confined to studying surfaces with no capability of going into the sample and studying subcellular structures or processes due to the very short range of the near-field region (E. Betzig et al. 1986; Huang, Bates, and Zhuang 2010).

The limitations of NSOM highlighted the need to develop super-resolution fluorescence methods based on the conventional far-field optics and the early nineties witnessed the debut of the first class of such methods (Huang, Bates, and Zhuang 2010; Hell et al. 2015). The main target of techniques belonging to this class is to produce sub-diffraction-limit features in the excitation pattern to sharpen the PSF, and hence they are sometimes referred to as "patterned excitation" methods (Huang, Bates, and Zhuang 2010). Among the most popular of these methods are stimulation emission depletion (STED) microscopy (Klar and Hell 1999) and saturated structured illumination microscopy (SSIM) (Gustafsson 2005). Both of these methods were able to achieve impressive resolutions of about 20 and 50 nm, respectively, but they still had their limitations. For instance, STED requires complex instrumentation and is only compatible with a narrow selection of fluorescent dyes, while SSIM may cause severe photobleaching of the fluorophores and requires post-processing computation to construct the image. Also, all of the patterned excitation methods typically depend on observing behaviors of

ensembles of molecules which are observed simultaneously. This leads to the loss of valuable information that can only be observed at the single molecule level (Rust, Bates, and Zhuang 2006; Huang, Bates, and Zhuang 2010; Schermelleh, Heintzmann, and Leonhardt 2010).

It wasn't until 2006 that a new revolution in the field of super-resolution fluorescence microscopy took place when a new class of methods collectively called single molecule localization microscopy (SMLM) was introduced to the scientific community.

1.2 Single-Molecule Localization Microscopy (SMLM)

The debut of SMLM came about with the release of three consecutive publications in 2006 by three different labs, each of which introduced a different super-resolution fluorescence microscopy method but in actuality, they were all based on the same single molecule localization principle (E. Betzig et al. 2006; Rust, Bates, and Zhuang 2006; Hess, Girirajan, and Mason 2006). The first of these methods was photoactivated localization microscopy (PALM), which was introduced by Eric Betzig's group and this was the breakthrough that landed him the Nobel prize in chemistry later in 2014 (E. Betzig et al. 2006; Chang 2014). The other two methods were stochastic optical reconstruction microscopy (STORM) and fluorescence photoactivation localization microscopy (FPALM) introduced by Michael Rust's and Samuel Hess's groups, respectively (Rust, Bates, and Zhuang 2006; Hess, Girirajan, and Mason 2006).

The developers of these methods realized that a fluorescence image of some biological structure is defined by the positions of the fluorophores mapping that structure. It is thus conceivable that a super-resolution image can be built by determining the position of each one of the fluorophores in the sample with high precision.

Another crucial realization is that even though the fluorescence signal coming from a single fluorophore is diffraction-limited, the precision in determining this fluorophore's position isn't. This is depicted by the following equation (Thompson, Larson, and Webb 2002):

$$\sigma = \sqrt{\frac{s^2 + a^2/12}{N_p}} \tag{1.1}$$

where σ is the localization error and it represents how precise the fluorophore localization is, s is the standard deviation of the PSF arising from the fluorophore, which is typically characterized by applying a two-dimensional Gaussian fit, a is the discretization size (camera pixel size) and N_p is the number of photons collected.

This equation tells us that the localization precision gets better the more photons we collect, so with enough photons, we can locate the centroid of the fluorophore with a precision much higher than the diffraction limit. This concept made it possible for single-molecule tracking of fluorescent probes with a spatial resolution of about 1 nm (Yildiz et al. 2003).

The biggest limitation though was that if multiple molecules were in close proximity, the localization precision decreases due to the overlapping signal. This hindered the widespread use of single-molecule imaging in studying biological samples since these samples typically contain a very high density of fluorophores. So, for a while single-molecule imaging was restricted to *in vitro* biophysical studies or to the unique *in vivo* cases where the fluorescent signal is sparse enough that single molecules can still be resolved.

SMLM methods made their huge breakthrough by circumventing that limitation through the use of photoswitchable fluorophores (G. H. Patterson and Lippincott-Schwartz 2002; Bates, Blosser, and Zhuang 2005). These fluorophores have the unique ability to switch between a

fluorescent and a dark state by the help of an activation wavelength different from the excitation one. This activation can be controlled to allow only a sparse subset of the fluorophores to be fluorescent at a given time, which makes high precision single-molecule localization of these otherwise diffraction-limited fluorophores possible. Eventually, these activated fluorophores will be irreversibly photobleached as a result of the continuous excitation giving space for a new subset of molecules to be activated and localized and hence, the spatially unresolvable signal becomes resolved in the time domain. After going through enough cycles of activation, localization and photobleaching, the final super-resolution image can be constructed by stacking all the localization events together (Fig 1.1a). It is worth mentioning here that SMLM methods are often used in conjunction with total internal reflection fluorescence (TIRF) microscopy. The reason is that the TIRF setup bends the incident light used to illuminate the sample at a sharp angle that forces it to totally reflect off the sample (Wegel et al. 2016). This allows for exciting only the fluorophores that are present in a thin layer of the sample called the 'evanescent wave' without exciting the fluorophores from the main bulk of the sample leading to a significantly improved signal to noise ratio and a much better localization precision (Rust, Bates, and Zhuang 2006; E. Betzig et al. 2006).

All SMLM methods mentioned above follow the same principle and they differ mainly in the type of fluorescent probe they were initially introduced with. PALM and FPALM both use photoactivatable or photoswitchable fluorescent proteins which can be expressed as fusion constructs with the protein of interest. PALM was initially introduced using EosFP (Eos fluorescent protein) that switches between a green and a red fluorescent states by the help of UV (ultraviolet) light activation while FPALM was introduced using PA-GFP (photoactivatable green fluorescent protein) that gets activated by UV light from a dark to a green fluorescent state

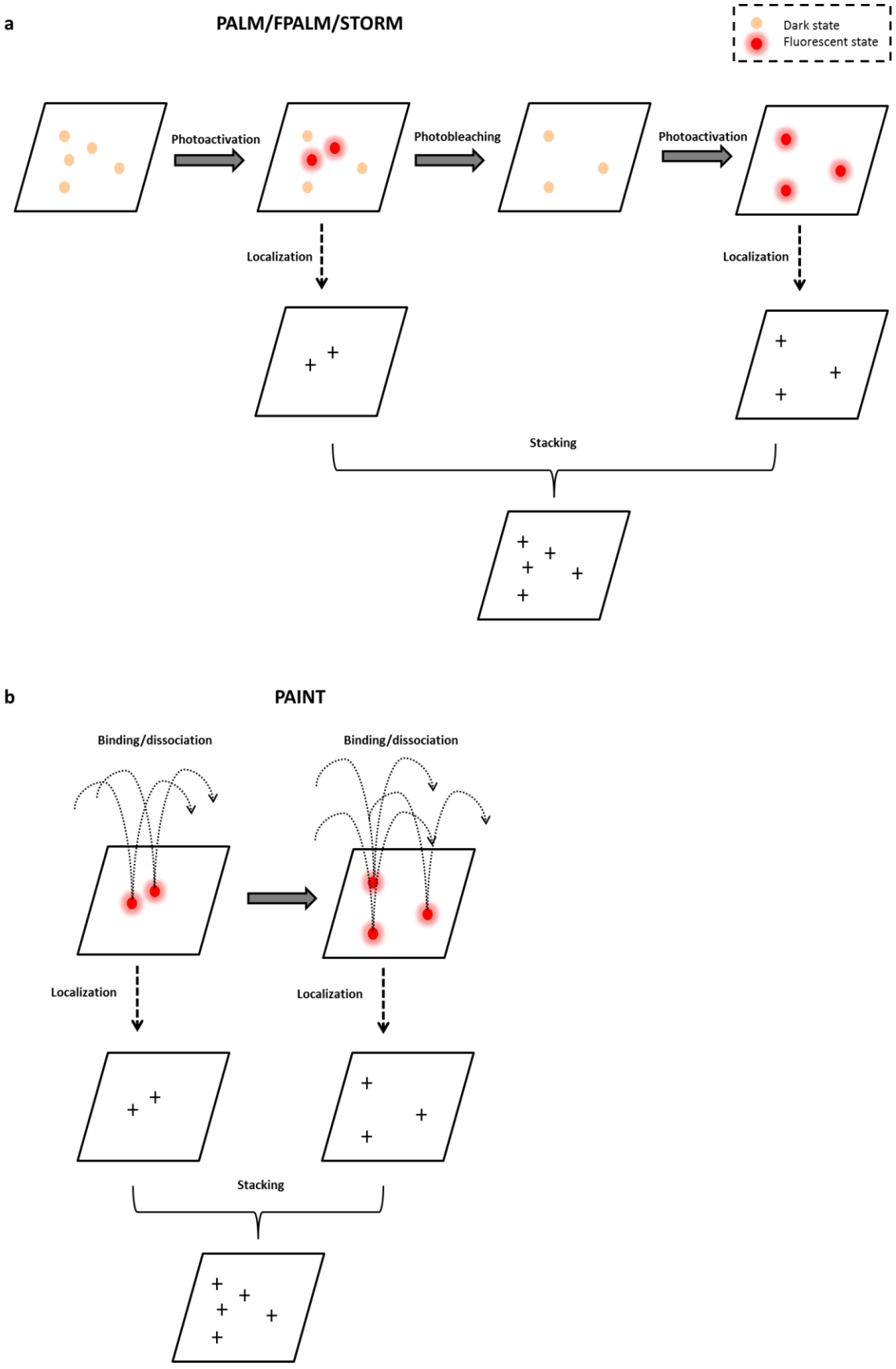

Fig 1.1. Description of different SMLM techniques. The process of acquiring the super-resolution image with PALM/FPALM/STORM (a) or PAINT (b).

(E. Betzig et al. 2006; Hess, Girirajan, and Mason 2006). On the other hand, STORM was introduced using the fluorescent dye Cy5, which can be attached to the protein of interest via antibodies and it switches between a dark and a fluorescent states with the help of deactivating red light and activating green light (Rust, Bates, and Zhuang 2006). Since fluorescent dyes typically emit more photons than fluorescent proteins, this difference gives STORM a slightly better resolution than PALM or FPALM, where STORM can reach a resolution of about 10 nm while PALM and FPALM can only reach a resolution of about 30 nm (Hell et al. 2015).

A slightly different variant of the previously mentioned techniques was also introduced in 2006 by Sharonov and Hochstrasser (Sharonov and Hochstrasser 2006). This technique is called PAINT (point accumulation for imaging in nanoscale topography) and it resembles the other SMLM methods in that it also depends on high precision single-molecule localization in rendering the final super-resolution image, but the difference lies in that instead of going through cycles of activation-localization-photobleaching of photoswitchable fluorophores, PAINT goes through cycles of binding-localization-dissociation of fluorescent probes diffuse in the solution. These probes keep targeting the feature to be imaged where their signal is recorded when they settle on that feature and become immobilized, then the signal is lost when the probe dissociates (Fig 1.1b). Reaching a low enough density of fluorescent probes to allow for high precision single-molecule localization is simply done by diluting the concentration of the probe in the imaging solution.

PAINT was initially introduced using Nile red probe that is non fluorescent in an aqueous medium and becomes fluorescent only in a hydrophobic environment. This probe was therefore used to image large lipid vesicles and supported lipid bilayers where it is diffuse in the solution and gets fluorescent when it binds to these lipid structures; but in principle, PAINT can be

carried out using virtually any fluorescent probe which is an advantage over other SMLM methods that require special probes (Sharonov and Hochstrasser 2006). Other advantages of PAINT include a diminished risk of photobleaching since using soluble fluorescent probes allows for an unlimited supply of new fresh probes, and the great potential for doing multi-color staining since a probe can be simply washed off and a new one added, which sets PAINT free of the limited wavelength choices available for other permanent labeling SMLM methods as PALM, STORM and FPALM.

Even though PAINT carried a lot of potential, it didn't get much traction until the work of Tai Kiuchi and coworkers in 2015 where they exploited the concept of PAINT by preparing a series of fluorescent probes that allowed them to image and overlay the spatial distributions of several cytoskeletal and focal adhesion proteins as actin, microtubules, plectin-1, PIPKIγ, paxillin and Src (Kiuchi et al. 2015).

SMLM methods easily became the most popular among the super-resolution fluorescence microscopy methods available to date given their many merits such as their readily achievable high resolution (10-30 nm), ability to explore single-molecule-level information even within high density structures, simplicity and convenience as they don't require special complicated instrumentation.

Nevertheless, SMLM methods still suffered from serious problems that are going to be discussed in the following section.

1.3 Problems with SMLM

Among the problems that face SMLM methods are: background noise that interferes with the single-molecule localization precision, the difficulty to do multicolor imaging due to the limited range of wavelengths available for photoactivation and excitation processes, and the poor temporal resolution that limits the use of SMLM in live imaging of rapid dynamic processes in the cell (G. Patterson et al. 2013).

In addition to the previous, there are two particularly challenging problems associated with SMLM methods that are the focus of this study. These problems are:

- The sample drift problem, and
- The measurement noise problem.

1.3.1 Sample drift problem

The process of constructing a super-resolution image using SMLM methods requires acquiring tens of thousands of frames where random subsets of single molecules are localized after either photoactivation in case of PALM/STORM/FPALM or binding of the diffuse probe in case of PAINT. During that extended imaging time, the stage where the sample is placed on is prone to drifting from the original position due to mechanical relaxation or temperature changes. This sample drift could be minimized but is generally unavoidable and the rate by which the sample drifts varies greatly based on the setup, but it can usually reach 1-10 nm/sec (Lee et al. 2012). This drift rate is insignificant for regular fluorescence microscopy with about 200 nm diffraction-limited resolution, but for SMLM methods with 10-30 nm resolutions, sample drift becomes a critical problem that might decrease the resolution of the constructed final image and distort the fine structures of the image (Fig. 1.2a).

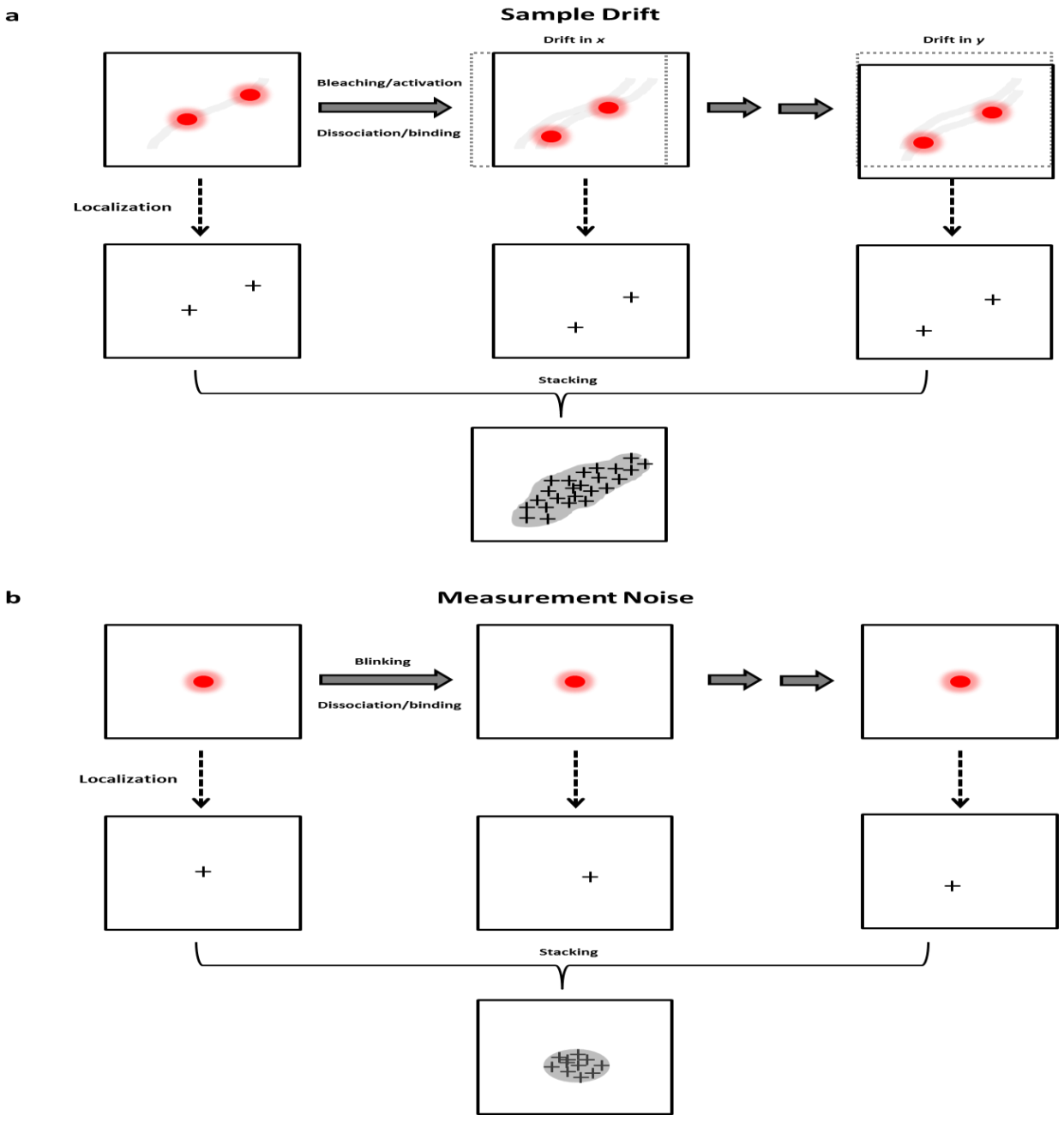

Fig. 1.2. Description of the sample drift and measurement noise problems. (a) The sample drift problem where the single-molecule data from the feature of interest is acquired sequentially through cycles of photoactivation/localization/photobleaching in case of PALM/STORM/FPALM or cycles of binding/localization/dissociation in case of PAINT. The drift in the *x* and *y* directions keeps changing the feature position and sequentially the newer localizations will be shifted leading to a distorted, lower resolution constructed image. (b) The measurement noise problem that manifests either by multiple detections of the same molecule undergoing blinking (PALM/STORM/FPALM), or by the dissociation and rebinding at the same position by diffuse probes (PAINT). In both cases, the localization uncertainty might give different localization coordinates every time the molecule reappears or rebinds leading to a blurred, lower resolution constructed image and the apparent cluster morphology of a single point source.

Many groups have tried to overcome this problem and they came up with different approaches to do that but generally, these approaches could be categorized into two groups:

1- Direct methods that try to measure the drift during the acquisition process.
2- Post-processing methods that try to estimate the drift from the acquired single-molecule data.

The first group of these methods uses hardware implementations to make the drift estimates. One way to do that is by modifying the optical setup to either take secondary bright-field images simultaneous to the fluorescent single-molecule images, or by adding active feedback equipment to stabilize the stage (Tang et al. 2014). Both methods work fine but they have the complication of adding extra equipment that might significantly increase the cost.

Another popular way of estimating the drift directly during acquisition is by adding fluorescent beads to the sample being imaged (Lee et al. 2012) to be used as fiduciary markers. These beads get imaged simultaneously with the sample and their tracks give the drift estimates. The limitations of this method include adding to the complexity of the experimental setup, the markers could interfere with the real single-molecule signal and they often themselves photobleach gradually, which would affect their localization precision and hence, the drift estimation.

The second group of methods doesn't apply any hardware modifications, but instead they are based on the image correlation analysis of the acquired single-molecule data (Geisler et al. 2012; Wang et al. 2014; Mlodzianoski et al. 2011). This is carried out by binning frames together to form substacks of the original full stack, then fixing one of these substacks as a reference while moving the others in the *x* and *y* directions till maximum correlation with the reference is

achieved. These *x* and *y* shifts represent the drift estimates between the substacks and the drift between individual frames is sequentially estimated by interpolation. It is worth noting that image correlation cannot be used to estimate the drift between individual frames directly because each frame doesn't carry enough single-molecule information for an accurate calculation of the correlation function.

Image correlation is a very simple and straight-forward method but the need to estimate the drift from substacks instead of individual frames results in the poor temporal resolution achieved by this method. Another problem that stems from this is that we don't have any prior knowledge on the optimal number of frames to be binned together to form these substacks. This number is crucial since binning too few frames would lead to erroneous estimation of drift while too many frames would worsen the temporal resolution even further (Geisler et al. 2012). Furthermore, while image correlation might work well with smooth drifts, it is likely to perform poorly if the drift has sudden large jumps caused by the build-up of mechanical stress.

The limitations of the methods available for sample drift estimation and correction highlighted the need for a better approach that would be able to estimate the drift at the single-frame temporal resolution and at the same time could be done using the single-molecule localization data without the need of adding any further hardware complexities to the experimental setup.

1.3.2 Measurement noise problem

As all experimental methods, SMLM suffers from measurement noise and one component of this noise is localization uncertainty. This uncertainty is depicted in Eq. 1.1 which states that the precision in localizing a molecule is primarily dependent on the number of photons collected

(Thompson, Larson, and Webb 2002). As a result, two features separated by a distance smaller than this localization uncertainty will not be resolved. The other component of measurement noise is counting error that results from multiple detections of the same molecule. Combining these two measurement noise errors typically leads to blurred features in the final super-resolution image (Fig. 1.2b).

In PALM/STORM/FPALM, multiple detections of the same target molecule take place through photoblinking, which is a naturally occurring phenomenon associated with fluorescent proteins and dyes where the fluorophore goes into a temporary dark state before it recovers to its bright state (Annibale et al. 2011). Current SMLM image rendering algorithms would treat a fluorophore undergoing multiple cycles of blinking and recovering as different fluorophores assigning a different two-dimensional Gaussian spot to each one of them. This wouldn't be a problem if we are absolutely confident about the fluorophore's location, but due to localization uncertainty, every time the fluorophore is detected it could be given offset coordinates from the previous detection. As a result, when the final super-resolution image is constructed by stacking these localizations together, features will appear blurred and randomly distributed molecules will have an 'apparent' cluster morphology (Fig. 1.2b). This latter realization allowed us to revisit a previously accepted notion that clustering is a dominant feature of membrane proteins as it pointed out that these clusters might not be true clusters (Garcia-parajo et al. 2014). Given how problematic it is, serious attempts have been undertaken to try and minimize photoblinking during acquisition of single-molecule images. Developing protocols that modify the buffer system used in the imaging medium are among these attempts (Vogelsang et al. 2008), but these can only reduce the effect of photoblinking without completely getting rid of it.

Photoblinking is a problem relevant to SMLM methods that depend on permanent labeling of the sample like PALM/STORM/FPALM, but the same principle of multiple detections applies to transient labeling SMLM methods as PAINT where the same target molecule could be detected by multiple binding events of the diffuse fluorescent probe. Once again, these different binding events to the same target could be detected with offset coordinates from the true target location leading to the same distorting effects discussed above (Fig. 1.2b).

Given that multiple detections of the same target molecule are unavoidable in all SMLM methods, we are urged to focus on alleviating the localization uncertainty problem. This effort should involve identifying the shortcomings of the current SMLM image rendering algorithms and developing better algorithms that can overcome these shortcomings and approach the problem in a more efficient way.

Besides localization precision, another recently appreciated factor that greatly influences the spatial resolution in SMLM is labeling density, which is defined as the number of detected fluorescent labels per unit of space of the sample (Shroff et al. 2008). In general, we need a sufficient amount of fluorescent labels per unit of space to efficiently highlight the underlying feature of interest otherwise, artifacts in the final super-resolution image will be observed causing continuous features to appear discontinuous and greatly diminishing the resolving power.

Appreciating its importance, attempts have been made to find a mathematical characterization of the effect of labeling density on spatial resolution in SMLM. One popular attempt takes advantage of the Nyquist sampling theorem and says that in order to reach a certain resolution d, we need to have a spatial labeling frequency f that is half of that desired resolution; i.e.: $f =$

$d/2$ (Shroff et al. 2008). For example, if we want to achieve a two-dimensional spatial resolution of 20 nm, we need to detect a fluorescent label every 10 nm, which translates to a labeling density of 10^4 per μm^2. Unfortunately, the heuristic nature of this approach begs for a more rigorous mathematical definition. Furthermore, it doesn't consider the heterogeneity of biological structures that might lead to disparate local densities.

In conclusion, it becomes clear that an algorithm that would efficiently characterize the effect of localization uncertainty on spatial resolution in SMLM needs also to consider the uncertainty associated with the labeling density and give it a more rigorous mathematical definition.

1.4 Statistical Methods

This thesis is centered on the use of novel statistical approaches to address the SMLM problems mentioned above and in this section, I will review the main statistical concepts that will be further explored in the following chapters.

1.4.1 Bayesian inference

One important aspect that differentiates Bayesian inference from classical statistical inference is that it treats the unknown quantity of interest as a random variable instead of an unknown constant. As a result, the task of a Bayesian statistician is to make inferences on the underlying probability distribution of that random variable. To fully grasp that concept let's assume we have a random variable of interest X and that we have observed some data Y related to that random variable. Using Bayes theorem, we can describe the conditional probabilities of x and y, which are values from the X and Y distributions as follows (Bayes and Price 1763):

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{1.2}$$

where $P(x|y)$ is referred to as the 'posterior distribution' of the random variable given the observed data and it is what we are trying to infer, $P(y|x)$ is the 'likelihood' of the observed data given some value of the random variable, and $P(x)$ is the 'prior distribution' of the random variable, which represents the prior belief about its distribution before observing any data. $P(y)$ is a normalization constant independent of the value of the random variable and it can often be conveniently dropped to give the proportionality relationship:

$$P(x|y) \propto P(y|x)P(x) \tag{1.3}$$

Eqs. 1.2 and 1.3 point out another unique feature of Bayesian inference, which is that it incorporates the prior knowledge with the accumulating data to update the posterior probability distribution.

Many methods are available to draw inferences about the posterior probability distribution and one very popular among those is the maximum a posteriori (MAP) estimation method. MAP's target is to estimate $\hat{x}$ that maximizes the posterior probability. This can be formulated mathematically as:

$$\hat{x}_{MAP} = \arg\max_{x} P(x|y) \tag{1.4}$$

and by substituting from Eq. 1.3 we get:

$$\hat{x}_{MAP} = \arg\max_{x} P(y|x)P(x) \tag{1.5}$$

In the context of the sample drift problem, x and y are substituted with the drift $\boldsymbol{d}$ and the observed single-molecule localization data $\boldsymbol{o}$, respectively, and Eq. 1.2 becomes:

$$P(\boldsymbol{d}|\boldsymbol{o}) = \frac{P(\boldsymbol{o}|\boldsymbol{d})P(\boldsymbol{d})}{P(\boldsymbol{o})} \tag{1.6}$$

Following the same logic in Eq. 1.3-1.5, we can get to a MAP estimation problem whose aim is to look for the drift values that would maximize the drift posterior probability distribution as follows:

$$\begin{aligned} \widehat{\boldsymbol{d}}_{MAP} &= \arg\max_{\boldsymbol{d}} P(\boldsymbol{d}|\boldsymbol{o}) \\ &= \arg\max_{\boldsymbol{d}} P(\boldsymbol{o}|\boldsymbol{d})\, P(\boldsymbol{d}) \end{aligned} \tag{1.7}$$

and this is the main problem we are trying to solve to estimate the sample drift as will be shown later in chapter 2.

Another class of inference methods is based on sampling from the posterior distribution. One of the simplest of these methods is called Gibbs sampling, which we use in approaching the measurement noise problem and it will be discussed in a later subsection.

1.4.2 Expectation maximization algorithm

The expectation maximization (EM) algorithm is an iterative method that can be used to compute MAP estimates and it is particularly efficient when the statistical model involves a latent variable (Dempster, Laird, and Rubin 1977).

Let's assume a statistical model that involves observed data Y, latent variable X and unknown parameters θ. If we want to estimate θ, we will need to find its value that would maximize $P(y|\theta)$. This probability distribution is termed 'likelihood' consistent with the notation we used before and hence, maximizing this term is called maximum likelihood estimation (MLE), which is very close to MAP except for the lack of the prior distribution term. This likelihood term needs to involve the latent variable to give the joint distribution:

$$P(y|\theta) = \sum_x P(y, x|\theta) \tag{1.8}$$

This joint distribution is often very difficult to compute directly given that we have two unknowns, θ and x, but EM circumvents that by breaking this problem into two more manageable ones. It starts by making some initial guess for θ, then goes through iterations of:

E-step: where the expectation value of the log-likelihood is computed based on the known $\hat{\theta}^i$:

$$Q\left(\theta, \hat{\theta}^i\right) = \mathbb{E}[\log P(y, x|\theta)|y, \hat{\theta}^i] \tag{1.9}$$

where $\mathbb{E}$ means expectation value and the logarithmic form makes computation easier. This step basically makes an update of the distribution of the latent variable.

M-step: where a new optimization is done on θ:

$$\hat{\theta}^{i+1} = \arg\max_{\theta} Q(\theta, \hat{\theta}^i) \tag{1.10}$$

where this new θ is guaranteed to be better than the older one (Dempster, Laird, and Rubin 1977).

Iterations of EM will go on until convergence to the best estimates of both x and θ, and this general workflow is depicted in Fig. 1.3.

Moving on to how EM algorithm fits in the thesis, we need to take a look back at our MAP estimation problem for the sample drift in Eq. 1.7, and notice that it is very difficult to solve directly but if we introduce the parameter $\boldsymbol{\theta}$, which represents in this case the final super-resolution image, the problem becomes:

$$\widehat{\boldsymbol{d}}_{MAP} = \arg\max_{\boldsymbol{d}} P(\boldsymbol{o}|\boldsymbol{d},\boldsymbol{\theta})P(\boldsymbol{d}), \tag{1.11}$$

which is a simpler problem to solve using the EM framework that will iteratively make better estimations for both $\boldsymbol{d}$ and $\boldsymbol{\theta}$, simultaneously, as will be shown in chapter 2.

1.5 Thesis Objectives

The main goal of this thesis is to find proper statistical approaches to address specific issues with SMLM techniques. The first of these is the sample drift issue that greatly distorts the final super-resolution image and is generally unavoidable. The current approaches that try to tackle this problem involve both hardware modifications and post-processing analysis methods, but unfortunately both approaches suffer from serious disadvantages that undermine their usefulness.

The other issue is the measurement noise problem, which is another unavoidable problem with SMLM techniques. Current image rendering algorithms fail to quantify the effect of this noise on the final super-resolution image and they even introduce unnecessary blurring to these images because they cannot correct for localization uncertainty. These shortcomings highlight the need for a better approach that can also have some rigorous characterization of the

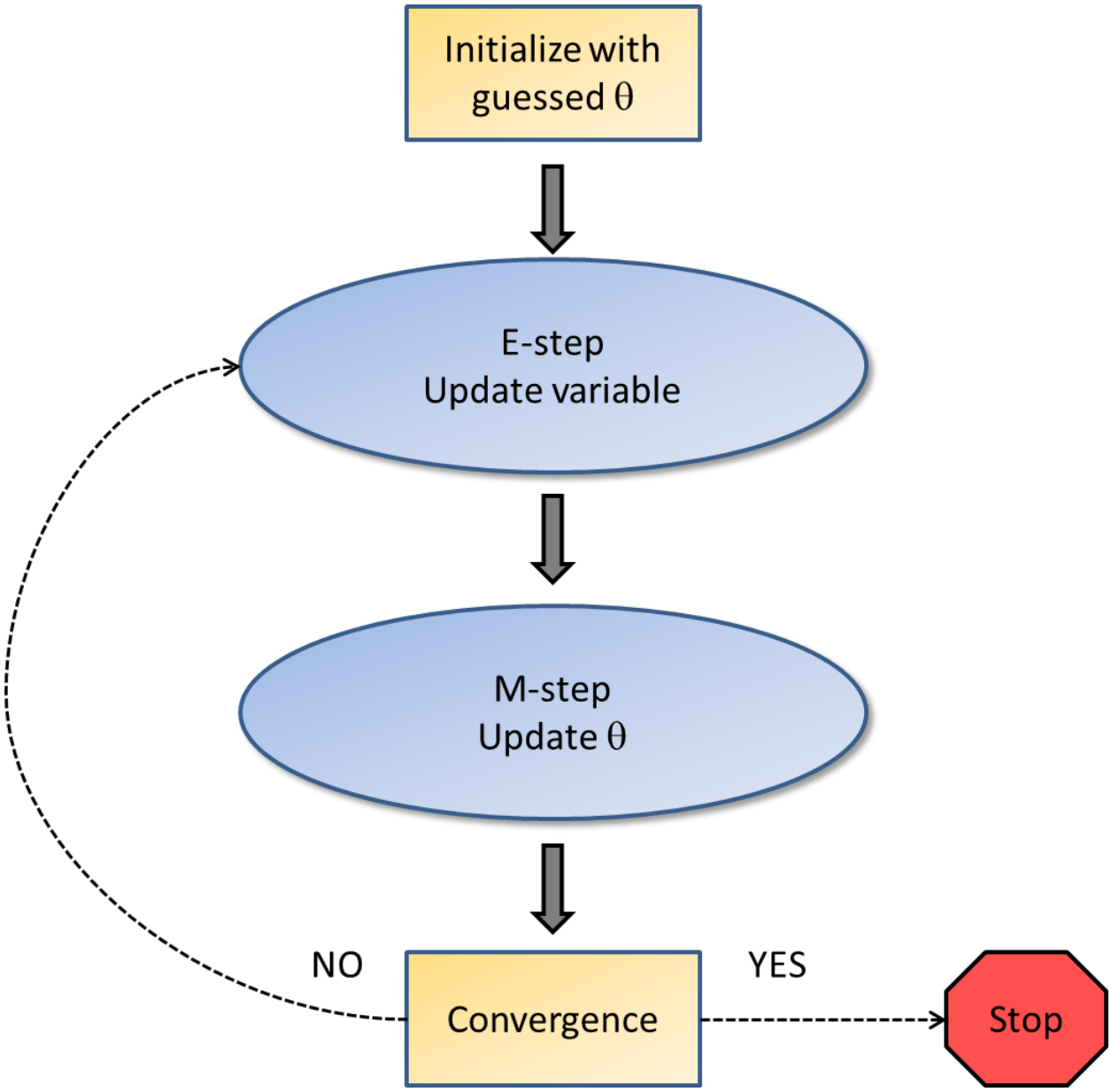

Fig 1.3. Workflow of the expectation maximization algorithm.

uncertainty associated with the effect of labeling density.

In chapter 2, I will demonstrate a statistical algorithm that applies MAP Bayesian inference to address the sample drift problem. I will validate the algorithm with simulated and experimental data then I will compare its performance with the famously used post-processing image correlation algorithm also using simulated and experimental data.

In chapter 3, I will demonstrate another statistical algorithm based on Gibbs sampling that addresses the measurement noise problem. The performance of the algorithm will be tested using simulated one- and two-dimensional data including different scenarios like filament-like structures and apparent and real clustering scenarios.

Chapter 2

Optimal Drift Correction for Super-resolution Localization Microscopy with Bayesian Inference

Ahmed Elmokadem and Ji Yu

This chapter was published in its present format in *Biophysical Journal*

Author contributions

AE performed experiments and analyzed data. JY designed the project and performed experiments. AE and JY contributed equally to the writing of the manuscript.

2.1 Abstract

Single-molecule localization based super-resolution microscopy requires accurate sample drift correction in order to achieve good results. Common approaches for drift compensation include using fiducial markers and direct drift estimation by image correlation. The former increases the experimental complexity and the latter estimates drift at a reduced temporal resolution. Here we present a new approach for drift correction based on the Bayesian statistical framework. The technique has the advantage of being able to calculate the drifts for every image frame of the dataset directly from the single-molecule coordinates. We present the theoretical foundation of the algorithm and an implementation that achieves significantly higher accuracy than image-correlation based estimations.

2.2 Introduction

Single-molecule localization based super-resolution microscopy has revolutionized optical microscopy field by pushing the spatial resolution to the scale of nanometers (E. Betzig et al. 2006; Hess, Girirajan, and Mason 2006; Huang et al. 2008). The remarkable improvement in spatial resolution comes at a cost of a more complicated imaging procedure: Instead of taking simple snapshots of the sample, tens of thousands of images are taken from the same sample, in which random subsets of the target molecules are "turned on" to be imaged and localized. The final "image" from the process is in the form of a histogram describing the frequency of the molecules being localized to certain spatial pixels. Sample drift during the data collection process can only be minimized, but is generally unavoidable. The popularity of this imaging method had resulted in extensive research on localization algorithms to process single-molecule imaging data and the efficiency and accuracy of various algorithms have been discussed in significant details (Sage et al. 2015; Small and Parthasarathy 2014). However, without accurate sample drift correction, the spatial resolution in the final reconstructed image will be poor even with the best localization accuracy.

Current sample drift correction techniques can be categorized into two groups: The first group attempts to directly measure the drift with hardware implementations. A popular technique is to add bright fiducial markers into the sample, which are co-imaged with the target molecules (Lee et al. 2012; E. Betzig et al. 2006). Other related techniques include the use of a secondary image of the sample (Bates et al. 2007; Tang et al. 2014). These techniques introduce extra complexities into the experimental procedure and are not always straightforward to implement, for example, fiducial markers often themselves photo-bleach gradually, which could result in shifting their centroid positions and thus errors in the drift measurements. A second group of

approaches is based on the idea of estimating drift directly from the single-molecule data using image-correlation (Geisler et al. 2012; Mlodzianoski et al. 2011; Wang et al. 2014; Huang et al. 2008). In general, the drift compensation of this type involves computing coarse super-resolution images based on sub-stacks of the total dataset, and computing sample drift of those sub-stacks using image-correlation. While simple to implement, the technique has the disadvantage that the drift is estimated at a coarse time-resolution. Furthermore, while the technique works well for drifts that are smooth, it could be problematic if mechanical creeps, which are sudden and large jumps in sample positions due to build-ups of mechanical strain, existed in the drift.

To offer a better approach for drift compensation, we treat it as a statistical inference problem. According to the Bayesian statistics framework, the estimation of the drift, $\boldsymbol{d}$, from the single-molecule dataset, $\boldsymbol{o}$, is the problem of obtaining a Maximum a posteriori (MAP) estimation:

$$\widehat{\boldsymbol{d}}_{MAP} = \arg\max_{\mathbf{d}} P(\boldsymbol{o}|\boldsymbol{d})P(\boldsymbol{d}) \tag{2.1}$$

Here, the dataset $\boldsymbol{o} \in \mathbb{N}^{W\times H\times N}$ is a three-dimensional matrix representing all N frames of individual super-resolution images. The size of each image is $W \times H$ pixels. For raw experimental data, the intensity values of each pixel can really only be either 0 or 1, depending on whether a molecule is detected at that pixel or not. But here we will deal with a slightly more general case, in which the intensity can be any natural numbers, i.e., 0, 1, 2 …. This allows us to deal with special cases where the raw frames were binned every few frames before drift inference, which is useful for reducing computational time for extremely large datasets. The drift, $\boldsymbol{d} \in \mathbb{Z}^{N\times 2}$, is a matrix representing time-dependent sample positions of all image frames. While this paper focuses on two-dimensional imaging, extension to three-dimensional case

should be straightforward. Furthermore, since the final constructed images are in pixelized form, it is unnecessary to compute drifts at unlimited resolution. Thus we model drift in the integer domain, assuming they are in the unit of pixels.

Importantly, the term $P(\boldsymbol{d})$ reflects our prior knowledge about the drift. Incorporating prior knowledge is a powerful way to achieve accurate inference from noisy data. In fact, the success of the image-correlation method is at least in part due to the fact that it implicitly incorporated a prior knowledge –the drift should be smooth in time. We explicitly model the prior distribution as a Markovian process (see methods); the displacements between adjacent frames were assumed to be approximately normally distributed, which favors smooth drifting traces.

Direct optimization based on Eq. (2.1) is, unfortunately, difficult, because the computation of the probability term involves high dimensional integration. However, we noted that if we had guessed a final super-resolution image, $\boldsymbol{\theta}$, the probability distribution of the drift could be estimated relatively easily according to the statistical modeling of $P(\boldsymbol{d}|\boldsymbol{o},\boldsymbol{\theta}) \sim P(\boldsymbol{o}|\boldsymbol{d},\boldsymbol{\theta})P(\boldsymbol{d})$. Based on this insight, we can iteratively make better estimation of $\boldsymbol{d}$ and $\boldsymbol{\theta}$, using the so-called expectation maximization (EM) algorithm (Dempster, Laird, and Rubin 1977), and finally reach co-optimization of both the most likely drift trace and the compensated super-resolution image $\boldsymbol{\theta}$. Details of the statistical model and derivation of the convergence formula are outlined in the methods section. We call our method BaSDI (Bayesian Sample Drift Inference).

2.3 Methods

2.3.1 BaSDI overview

Eq. (2.1) is difficult to solve directly. However, we note that the difficulty is because we do not know anything about how the molecules are spatially distributed in the sample. In other words, we don't have a good estimation of the true final super-resolution image $\boldsymbol{\theta} \in \mathbb{R}^{W\times H}$, where θ_{ij} is proportional to the molecular density at the spatial coordinate (i, j). For convenience, we can make sure that $\boldsymbol{\theta}$ is normalized:

$$\sum_{i,j} \theta_{ij} = 1 \tag{2.2}$$

Had we known the value of $\boldsymbol{\theta}$, the conditional probability $P(\boldsymbol{d}|\boldsymbol{o},\boldsymbol{\theta}) \sim P(\boldsymbol{o}|\boldsymbol{d},\boldsymbol{\theta})P(\boldsymbol{d})$ can be expressed in relatively simple analytical forms. To see that, we note first that since $\boldsymbol{\theta}$ is normalized, the probability of observing a single localization event at pixel location (i, j) is simply θ_{ij}. In a single image frame $\boldsymbol{f} \in \mathbb{N}^{W\times H}$, there are multiple molecules detected at various coordinates. The joint probability of observing all these localization events is therefore:

$$P(\boldsymbol{f}|\boldsymbol{\theta}) = (\|\boldsymbol{f}\|_1)! \cdot \prod_{i,j} \left(\theta_{i,j}\right)^{f_{i,j}}, \tag{2.3}$$

where $\|\boldsymbol{f}\|_1$ denotes the total number of localization events in $\boldsymbol{f}$, and the factorial term is to account for the permutations of all the sequences of the molecules. Nevertheless, the expression ignores the effect of drift. If the current image frame is captured with a drifted sample position (x, y), then we need to shift the indices of the observed image. Furthermore, it is often convenient to compute probability distribution in log terms:

$$\log P(\boldsymbol{f}|\boldsymbol{\theta}, x, y) \approx \log(\|\boldsymbol{f}\|_1)! + \sum_{i,j} f_{i-x,j-y} \log \theta_{i,j} \tag{2.4}$$

Here, we changed the equal sign to an approximate sign because the index shifting may have moved some of the observed localization events outside the reference frame, altering the total number of localization events considered. In practice, this approximation should not introduce large errors if we simply add an empty border of certain amount of pixels, *s*, to the images. Then, as long as the maximum drift does not exceed *s*, we can still perform the computation to obtain reasonable values of the posteriori distribution. In the real imaging experiment, data with too much drift (i.e. > s) is going to be problematic anyway. Therefore the approximate approach should not introduce extra limitations.

Finally, the likelihood of observing the whole dataset is simply the joint probability of all individual image frames:

$$\log P(\boldsymbol{o}|\boldsymbol{d},\boldsymbol{\theta}) \approx \sum_{k}\sum_{i,j} o_{i',j',k} \log \theta_{i,j} + constant \tag{2.5}$$

where

$$i' = i - d_{k,1}; \quad j' = j - d_{k,2} \tag{2.6}$$

Here we dropped the explicit expression of the permutation terms because they are independent of the parameters we are trying to optimize.

We propose that the optimization can be realized with an iterative computational algorithm based on the framework of expectation maximization (EM) (Dempster, Laird, and Rubin 1977):

E-Step: Given a guessed $\boldsymbol{\theta}$ matrix, compute the conditional distribution of $P\left(\boldsymbol{d}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[t]}\right)$, that is, what is the likelihood of certain drift traces if we already guessed the final super-

resolution image. A good initial guess for the $\boldsymbol{\theta}^{[0]}$ would simply be the summed image of all frames without drift correction.

M-Step: Based on the computed distribution of $P\left(\boldsymbol{d}\middle|\boldsymbol{o},\boldsymbol{\theta}^{[i]}\right)$, perform a new optimization of $\boldsymbol{\theta}$:

$$\widehat{\boldsymbol{\theta}}^{[i+1]} = \arg\max_{\theta} \mathbb{E}_{\boldsymbol{d}\in\boldsymbol{D}}\left[\log P(\boldsymbol{o}|\boldsymbol{d},\boldsymbol{\theta})\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right] \tag{2.7}$$

where $\mathbb{E}$ means the expectation value, and $\boldsymbol{D}$ is the set of all possible configurations of $\boldsymbol{d}$. It can be proven (Dempster, Laird, and Rubin 1977) that the new estimation $\widehat{\boldsymbol{\theta}}^{[i+1]}$ is guaranteed to be a better one than the older estimation, $\widehat{\boldsymbol{\theta}}^{[i]}$.

To perform optimization shown in Eq. (2.7), we first expand it by plugging in the result from Eq. 2.5.

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{[i+1]} &= \arg\max_{\theta} \mathbb{E}_{\boldsymbol{d}\in\boldsymbol{D}}\left[\sum_k \sum_{i,j} o_{i',j',k} \log\theta_{i,j} + constant\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right] \\ &= \arg\max_{\theta} \mathbb{E}_{\boldsymbol{d}\in\boldsymbol{D}}\left[\sum_k \sum_{i,j} o_{i',j',k} \log\theta_{i,j}\ \middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right]\end{aligned} \tag{2.8}$$

where the definitions of *i'* and *j*' are the same as in Eq 2.6. Next we shift the summation over image frames out of the expectation function, and express the expectation function as explicit summations:

$$\begin{aligned}\widehat{\boldsymbol{\theta}}^{[i+1]} &= \arg\max_{\theta} \sum_{\mathrm{k}} \mathbb{E}_{\boldsymbol{d}\in\boldsymbol{D}}\left[\sum_{\mathrm{i,j}} o_{i',j',k} \log\theta_{i,j}\ \middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right] \\ &= \arg\max_{\theta} \sum_k \sum_{d_{k,1},d_{k,2}} \sum_{i,j} P\left(d_{k,1},d_{k,2}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right) o_{i',j',k} \log\theta_{ij}\end{aligned} \tag{2.9}$$

Note that the optimization is constrained by the condition specified in Eq (2.2). To obtain the maximum value of this function, we use the standard Lagrangian optimization technique, by solving for the root of the derivative of the target function in addition to a Lagrangian term:

$$\frac{\partial}{\partial \theta_{ij}}\left[\sum_{k}\sum_{d_{k,1},d_{k,2}}\sum_{i,j} P\left(d_{k,1},d_{k,2}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right) o_{i',j',k}\log\theta_{ij} + \lambda\left(\sum_{ij}\theta_{ij}-1\right)\right]=0 \tag{2.10}$$

$$\lambda+\theta_{ij}^{-1}\sum_{k}\sum_{d_{k,1},d_{k,2}} P\left(d_{k,1},d_{k,2}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right)o_{i',j',k}=0$$

And it's easy to see from there that

$$\hat{\theta}_{ij}^{[i+1]}\propto\sum_{k}\sum_{d_{k,1},d_{k,2}} P\left(d_{k,1},d_{k,2}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right)o_{i',j',k} \tag{2.11}$$

Eq. 2.11 now allows us to compute for the M-step. The E and M steps iterate until the algorithm converges. The most updated source code for the implementation is available to download from https://github.com/jiyuuchc/BaSDI/releases/ .

2.3.2 Prior distribution for the drift *d*

In order to apply the EM algorithm, we need to compute $P\left(d_{k,1},d_{k,2}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right)$, which is a marginal probability that can be calculated by integrating $P\left(\boldsymbol{d}\middle|\boldsymbol{o},\widehat{\boldsymbol{\theta}}^{[i]}\right)$ over all frames that is not *k*. Note that this integration depends on the choice of the prior probability $P(\boldsymbol{d})$. However brute-force integration would be typically too time-consuming to be practical, thus the prior probability distribution needs to be designed carefully to satisfy two requirements: (a) it needs to enforce

some level of smoothness in the drift traces so as to agree with our physical intuitions about the drift, and (*b*) it should allow easier computation of the marginal probability. In our case, we use a simple function based on the random walk model where the stepping probability distribution is a truncated Gaussian distribution:

$$\begin{aligned} P(\boldsymbol{d}) &= P(d_{11})P(d_{12})\prod_{k=2}^{N} P\left(d_{k,1}, d_{k,2} \middle| d_{k-1,1}, d_{k-1,2}\right) \\ &= P(d_{11})P(d_{12})\prod_{k=2}^{N} t\left(d_{k,1} - d_{k-1,1}, d_{k,2} - d_{k-1,2}\right) \end{aligned} \tag{2.12}$$

$$t(\delta x, \delta y) = \mathcal{N}(\delta \mathrm{x}; 0, \sigma^2)\mathcal{N}(\delta \mathrm{y}; 0, \sigma^2) + \epsilon; \quad \delta \mathrm{x} < s, \delta \mathrm{y} < s \tag{2.13}$$

where $\mathcal{N}$ denotes normal distribution and σ^2 is a hyper parameter corresponding to the speed of the drift. The value $\epsilon > 0$ is used to account for the small probability of a "creep" – rare, sudden jumps with larger amplitude – in the system. We kept ϵ to be $1/N$ for all our calculations in this paper.

Because of the Markovian characteristics of $P(\boldsymbol{d})$, the marginal probability can be computed using the well-known forward-backward algorithm, which is much more efficient than the brute-force integration.

2.3.3 Forward-Backward Algorithm

Before delving into the forward-backward algorithm, we need to first explain what a "Markov process" is. A Markov process describes states that are evolving in time in which the future evolution has a probabilistic dependence on the past. The special thing about this dependence in

a Markovian process is that any future state is only dependent on the current state without having to keep track of all the past events (Blitzstein and Hwang 2015). Assuming a chain of states $X_1, \dots, X_n$, we can mathematically describe the Markovian dependencies between these states as follows:

$$P(x_k|x_1, x_2, \dots, x_{k-1}) = P(x_k|x_{k-1}) \tag{2.14}$$

Fig. 2.1 shows a special case of Markovian processes where the chain of states X are 'hidden', i.e.: cannot be observed directly but rather through indirect observations $Y_1, \dots, Y_n$. This special case is called a hidden Markov model (HMM) and it has a wide range of applications from speech recognition to DNA sequence analysis (Blitzstein and Hwang 2015).

The forward-backward algorithm is a perfect tool to draw inferences about Markovian models. More specifically, it is very efficient in computing $P(x_k|y_1, \dots, y_n)$, which is the marginal probability of a specific state given all observed data. It does that by computing two functions: $\alpha_k(x_k)$ and $\beta_k(x_k)$, where the α function is computed from the first to the last state while the β function is computed from the last to the first state as follows:

$$\alpha_1(x_1) = P(x_1, y_1)$$

$$\alpha_k(x_k) = P(y_k|x_k) \sum_{x_{k-1}} \alpha_{k-1}(x_{k-1}) P(x_k|x_{k-1}) \tag{2.15}$$

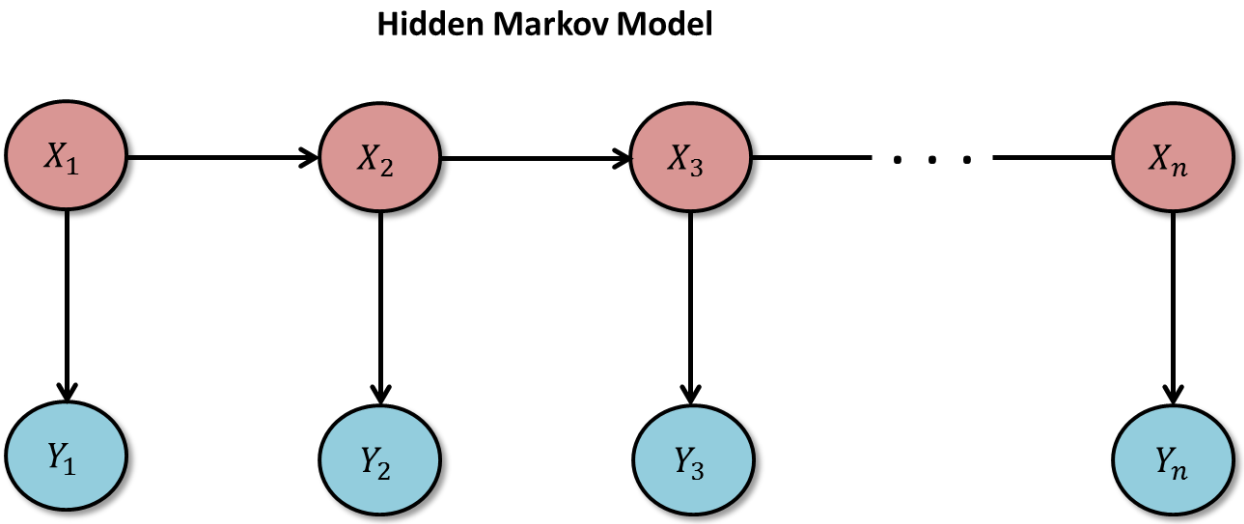

Fig. 2.1. Generic example of a hidden Markov model.

$$\beta_n(x_n) = 1$$

$$\beta_k(x_k) = \sum_{x_{k+1}} \beta_{k+1}(x_{k+1}) P(y_{k+1}|x_{k+1}) P(x_k|x_{k+1}) \tag{2.16}$$

where $\alpha_1(x_1)$ and $\beta_n(x_n)$ are the initialization states, $P(y_k|x_k)$ and $P(y_{k+1}|x_{k+1})$ are called the emission or observation probabilities and $P(x_k|x_{k-1})$ and $P(x_k|x_{k+1})$ are called the transition probabilities. Notice that the forward-backward algorithm owes its efficiency to its recursive nature where every computation of α and β makes use of the previous computation of these functions.

Finally, the marginal probabilities can be simply computed as:

$$P(x_k|y_1, \dots, y_n) \propto \alpha_k(x_k)\beta_k(x_k) \tag{2.17}$$

The sample drift can be thought of as a Markovian process since it represents an evolution in time where the future state only depends on the current one. This property enables us to use the efficient forward-backward algorithm in computing the marginal probabilities $P(\boldsymbol{d}_k|\boldsymbol{o}, \boldsymbol{\theta})$, which are needed by the EM algorithm to carry out the drift optimization. This can be done as follows:

$$P\left(d_{k,1}, d_{k,2}\middle|\boldsymbol{o}, \widehat{\boldsymbol{\theta}}^{[i]}\right) \sim \alpha(x, y, k)\beta(\mathrm{x}, \mathrm{y}, \mathrm{k}) \tag{2.18}$$

where both the α and the β functions are calculated iteratively. The α values are computed from the first frame to the last frame:

$$\begin{aligned} &\alpha(x, y, 1) \sim P\left(\boldsymbol{f}_1\middle|\widehat{\boldsymbol{\theta}}^{(i)}, x, y\right) \\ &\alpha(x, y, k) \sim P\left(\boldsymbol{f}_k\middle|\widehat{\boldsymbol{\theta}}^{(i)}, x, y\right) \sum_{\delta x, \delta y} \alpha(x, y, k-1) t(x - \delta x, y - \delta y) \end{aligned} \tag{2.19}$$

and the β values are calculated in reverse, starting from the last image frame:

$$\beta(x, y, N) = 1$$
$$\beta(x, y, k) \sim \sum_{\delta x, \delta y} P\left(\boldsymbol{f}_k \middle| \widehat{\boldsymbol{\theta}}^{(i)}, \delta x, \delta y\right) \beta(x, y, k+1) t(\delta x - x, \delta y - y) \quad (2.20)$$

In both Eq. 2.19 and 2.20, $\boldsymbol{f}_k$ denotes the observed data in the *k*-th image frame. Its probability calculation follows Eq. 2.4.

2.3.4 Efficiency of convergence

Although EM algorithm guarantees the increase of the likelihood with each iteration, it does not guarantee convergence to the global maximum (i.e., it can converge to a local maximum). We found that while the algorithm performs very efficiently when the estimated $\widehat{\boldsymbol{\theta}}^{(i)}$ image is smooth across most pixels, it tends to be trapped at a local maximum when $\widehat{\boldsymbol{\theta}}^{(i)}$ contains many pixels of zeros values. This suggests that for a dataset of low sampling rate – i.e. when the total number of detected molecules is too low to construct a smooth final PALM image – it is more efficient to compute $\widehat{\boldsymbol{\theta}}^{(i)}$ at lower resolution (or use a larger pixel size), which has the effect of smoothing out the global search space function allowing more efficient convergence to the global optimum. To accommodate all dataset scenarios, we designed the EM iterations to follow a multi-rounds optimization schedule: The initial rounds reduce the effective resolution of $\widehat{\boldsymbol{\theta}}^{(i)}$ by applying a smoothing filter. The size of the filter is gradually reduced in the later rounds, which iteratively refines the drift estimation until at the final round, no smoothing filter is applied.

The default optimization schedule we implemented in the accompanying software performs very well for all varieties of datasets and had yet to fail once in all the simulated datasets we have generated for testing so far. Alternatively, to save computational time, one could skip earlier rounds of iteration if the dataset sampling rate is high, and can skip later rounds if sampling rate is low.

2.3.5 Simulation and algorithm validation

PALM datasets for algorithm validation were generated with Monte Carlo simulation. To do that, we first produce a ground truth image with certain characteristic "features". Molecules of interest are assumed to be concentrated within these "features" and much less abundant outside (the background). The values of the ground truth image were assumed to represent the concentration of the molecules within the area of each image pixel. To generate individual frames of localization images, we simply produced Bernoulli samples for each pixel position, for which the Bernoulli probabilities are assumed to be proportional to intensity of the ground truth image. The Bernoulli sampling ensures that each pixel can have no more than one molecule in the localization image, similar to real experiments where molecules are localized individually. After a stack of multiple images were generated, each image was shifted according to a drift trace also produced from Monte Carlo simulation (see below). The Bernoulli probabilities are assumed to be unchanged for all image frames. It should be noted that based on this generation process, all images were produced independent of each other. In other words, in our simulation the detected molecules will generally not persist over multiple image frames. In real experiments, the "switching off" of a single molecule is not instantaneous, thus signal from a single molecule sometimes persists over consecutive frames. This type of temporally-correlated data provides extra information about the drift represented in the lateral translocations of those

individual molecules. However, experimentally it is also desired to "switch off" any detected molecules as quickly as possible in order to increase data acquisition speed. Therefore, we opted to not utilize this extra temporal information and the simulation is designed to test the validity of the algorithm in the most general case possible.

Drifts were generated according to a random-walk model, i.e. $d_{t+1} = d_t + \mathcal{N}(0, \sigma^2)$. Values of σ^2 were chosen randomly from 0.05 – 0.5, unless otherwise specified, to simulate various rates of drift, although its value does not change within the same drift trace. To generate "persistent" drift traces, we model the velocity as a random walk: $\dot{d_{t+1}} = \dot{d_t} + \mathcal{N}(0, \sigma_v^2)$. Regardless of the ways the drifts were generated, for all simulations the inferences by BaSDI were performed with the fixed prior hyperparameter of $\sigma^2 = 0.1$.

For image correlation analysis, the data was grouped into various numbers (5-20) of substacks and the best groupings were empirically chosen by comparing the results to the known input drift traces. This is of course impossible for real experimental data, where the "true" drift is unknown. Thus for real-experimental data, we used the grouping that performed best based on our simulations.

2.3.6 BaSDI application to experimental PALM microscopy data

The human Crk cDNA sequence was a gift from Dr. Bruce Mayer (University of Connecticut Health Center) and subcloned into a mammalian expression vector carrying mEos3, producing the mEos3-Crk fusion construct. MCF-7 or MEF cells were transfected with 500 ng of DNA using 1.5 μL of Lipofectamine 2000 (Invitrogen) for 6 h, before being replated on plasma-cleaned glass-bottom dishes (MatTek) overnight. Cells were then fixed with 4% paraformaldehyde, washed twice with PBS and imaged with PALM with 561-nm total internal

reflection laser illumination and 405-nm photo-activation. Data acquisition was essentially the same as previously described (Tatavarty et al. 2009). Image processing and single-molecule localization were performed using the octane software (Tatavarty et al. 2009). The localization data were then used as input for BaSDI, which computed the optimized drift traces as well as the final constructed super-resolution image in an unsupervised manner.

2.4 Results and Discussion

2.4.1 Algorithm validation with simulated image data.

To validate BaSDI algorithm, we used Monte Carlo simulation to generate various testing datasets and tested whether BaSDI correctly estimated drift from those datasets (Fig 2.2). The ground-truth image is designed so that the features (central squares) occupy an area of 4900 pixels^2 in a total image of 250000 pixels^2 (~2% of total area). We then assigned a 10 times more probability of finding a molecule in a feature pixel than in a background pixel. As a result, about 20% of the detected molecules will appear within the features while the remaining 80% will be in the background. To mimic the localization microscopy data, frame-by-frame molecular coordinates were simulated by random sampling single-molecule coordinates from a ground-truth image (Fig 2.2a), assuming 50 molecules were detected on average for each image and 1000 total images were acquired. Each image was shifted according to a 2D drift trace following the random walk model to mimic experimental sample drifts. We then reconstructed the super-resolution image (Fig 2.2b) either without any drift correction, or performing drift correction by feeding the simulated imaging data to BaSDI. To estimate statistical error, the test was repeated 30 times by resampling the same ground truth image to produce new randomized datasets. It can

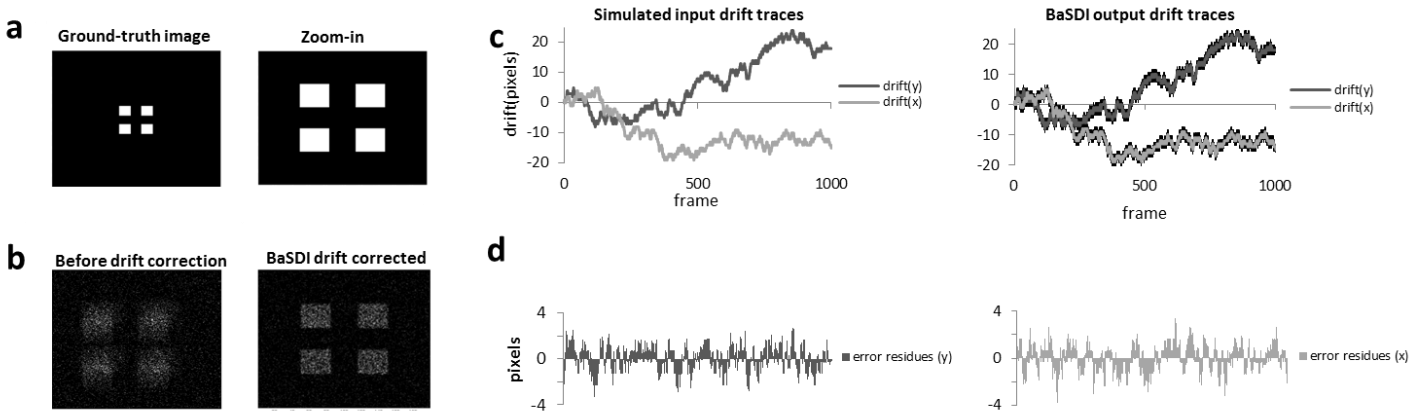

Fig 2.2. Assessing BaSDI's performance with simulated single-molecule localization data. (a) The ground-truth image (left) for simulation and a zoom-in view of its key feature (four squares at the center) (right). (b) Simulated PALM images before (left) and after (right) drift correction by BaSDI. The input drift traces were simulated according to a displacement random walk model. Single-molecule coordinates are generated by randomly sampling the ground true image in (a), assuming detection of 50 molecules per frame in average and 1000 frames per dataset. (c) The input drift traces (left) and the corresponding BaSDI output drift traces (right) in the y (vertical) and the x (horizontal) directions. Statistical noises in the BaSDI calculation, represented by the line with in the right panel, were estimated by rerunning the simulation and the corresponding drift recalculation 30 times. (d) Error residues obtained from the differences between the traces in (c) in the y and the x directions.

be shown that the reconstructed PALM image without drift compensation is distorted and differs significantly from the ground truth image (Fig 2.2b). However, BaSDI was able to correctly estimate the drift traces (Fig 2.2c&d), and after drift correction, produced a high resolution image matching the ground truth (Fig 2.2b).

Real biological samples can exhibit many different types of spatial features. For example, cytoskeletal structures are made of mostly linear elements, and cell surface receptor clusters are "spotty" structures. Thus we next tested how the underlying image features affect BaSDI performance (Fig 2.3). Besides the "blocky" image as shown in Fig 2.2, we also generated two new ground-truth images; one with predominately linear features (Fig 2.3c) and another with spotty features (Fig 2.3e). All ground truth images have the same integrated signal-to-background ratio, i.e., 20% of the total molecules residing within the "features". We found, again with the simulated localization image data, that BaSDI was able to significantly improve the quality of the reconstructed image by correcting the drift in the dataset. To quantitatively compare the performances in all three cases, we used Pearson correlation coefficient to evaluate the similarity between the reconstructed images with the ground-truth image (Fig 2.3b, d & f). Unsurprisingly, the uncorrected images had very low correlation with the ground truth images and BaSDI correction significantly improved the results. Interestingly, since we knew the exact drift used to produce the data, we could also construct images with perfect drift correction (Fig 2.3a, c & e), and compute their correlation with the ground truths. Even with no residue drift, these images do not exhibit perfect correlation due to noise rising from sampling. More importantly, we observe no significant differences between their correlation values with the values of the BaSDI-corrected images, indicating that in terms of image quality, the BaSDI corrected images are not statistically distinguishable from images with no drift at all in these

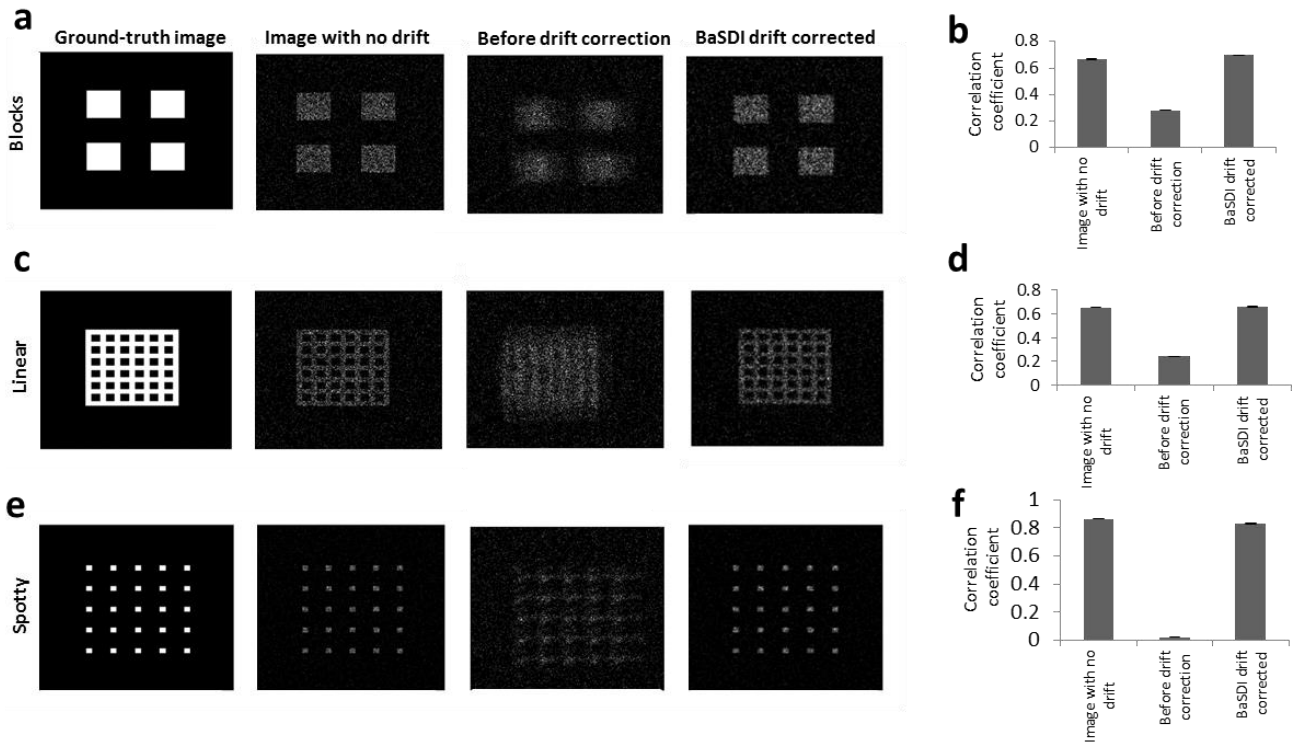

Fig 2.3. Assessing BaSDI's performance for drift correction given different PALM image types. Different ground-truth images for simulation (a) blocks, (c) linear, and (e) spotty images are shown on left followed by the PALM images without any drift, images before drift correction, and images after drift correction by BaSDI. (b), (d) and (f) show the corresponding correlation coefficients of each of the PALM images with the ground truth reference images. Error bars are s.e.m, n = 30.

simulations. In comparing results from different ground truth images, we find that the best results were obtained from the "spotty" images, while the "blocky" images have the worst results. This is also not surprising, as intuitively one would expect that higher spatial frequency would allow higher confidence in determining the drift. Since the "blocky" images represent a harder problem to solve, all the following simulation tests will be performed on this particular ground truth image.

In our statistical model, we assumed a Markovian-type prior distribution function for the drift. However, different microscopes may also have very different drift characteristics due to differences in their mechanical properties. Therefore, we further tested whether BaSDI algorithm can handle drift traces generated with other drift models (Fig 2.4). In addition to the random walk model (Fig 2.4a) similar to the data shown earlier, we further examined two additional drift generation models: a model in which the drift velocity (instead of drift position) is undergoing random walk (Fig 2.4b), which results in a more "persistent" drift trace, and a model in which random mechanical "creeps" (sudden jumps of large amplitude due to release of built-up stress) were introduced in a random-walk drift (Fig 2.4c). We found that in all of these scenarios BaSDI was able to estimate the drifts incorporated into the dataset. Still, realistic drift characteristics can only be obtained by examining the drift from a real microscope. Therefore, we obtained real experimental drift traces by tracking fluorescent beads from our own PALM microscope (Fig 2.4d), and produced simulated PALM datasets using these experimentally obtained drift traces. The experimentally obtained drift traces are similar in absolute amplitude to the simulated traces in previous figures. We found that BaSDI also successfully estimated the drift from this dataset (Fig 2.4d). To quantify BaSDI's performance in estimating drift traces from all different drift input scenarios, we calculated the root-mean-square (RMS) error of the estimated drifts (Fig

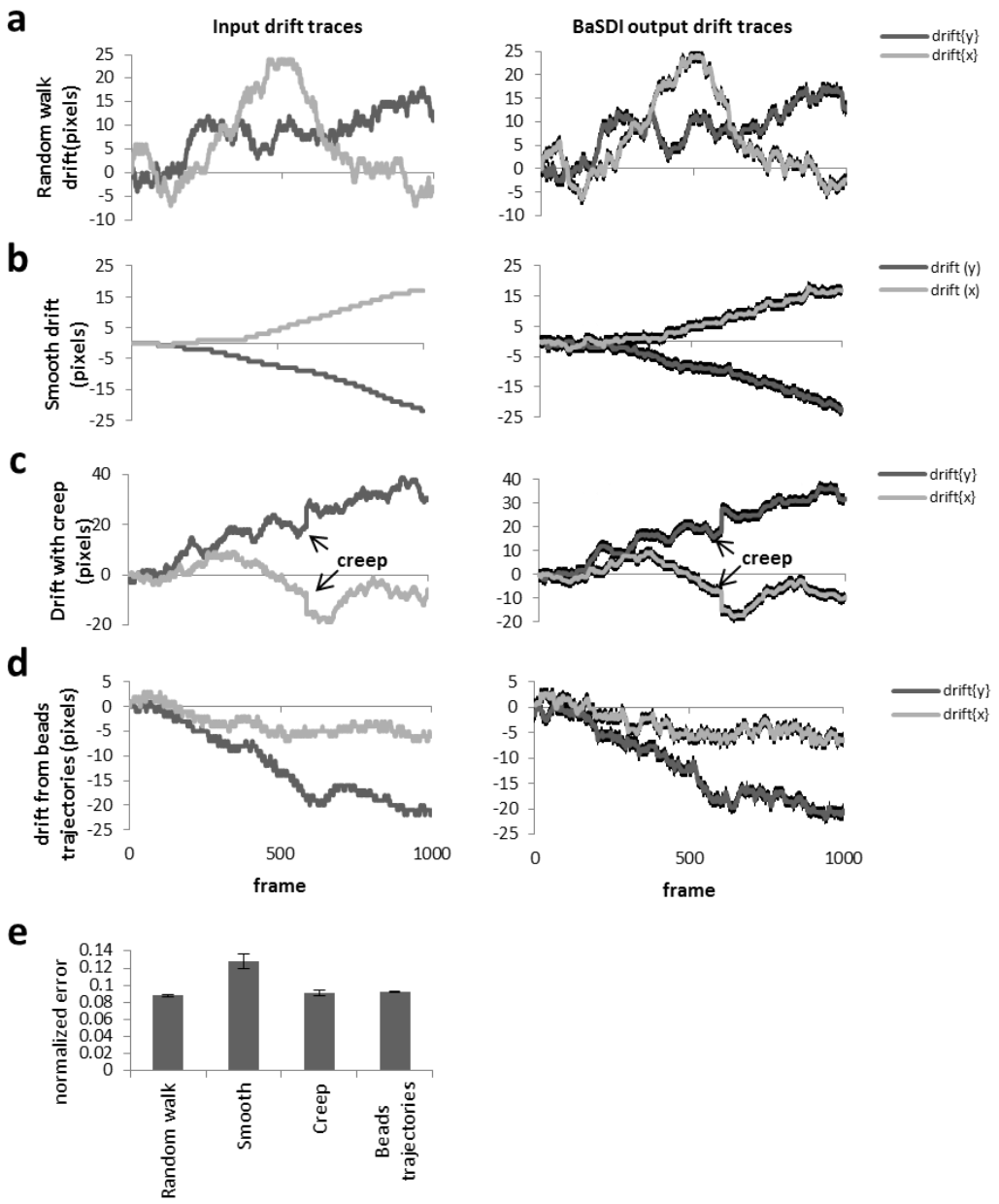

Fig 2.4. Assessing BaSDI's drift estimation of different drift types. BaSDI has been tested against (a) random walk drift, (b) smooth drift, (c) drift with mechanical creep, or (d) estimated drift from fluorescent beads trajectories. The input drift traces in both, the y (vertical) and x (horizontal) directions are shown on left panels on (a-d) while the BaSDI estimated output drift traces for each type are shown on right. (e) Shows the normalized error for BaSDI estimates against each type of input drift. Error bars in (a-d) are standard deviation while error bars in (e) are s.e.m, n = 30.

2.4e). We found no significant differences in the RMS errors of all three different drift models, suggesting similar BaSDI performances in all cases.

2.4.2 Algorithm validation with experimental image data

A true test of the algorithm has to be based on its performance on real PALM experimental data. We performed PALM imaging experiments on MCF-7 cells expressing mEos3-Crk. Crk is an adaptor protein molecule that can induce cell transformation (Matsuda et al. 1992; Zheng et al. 2010). It contains a Src homology 2 (SH2) domain that allows it to bind to various tyrosine-phosphorylated proteins in cells. In mammalian cells, Crk is known to localize to focal adhesions (FAs) (Zheng et al. 2010), which typically contain concentrated phosphotyrosines, but also to cell surface receptor tyrosine kinases (Abassi and Vuori 2002), which can be phosphorylated as well. The competition between the two as well as the difference in phosphorylation dynamics between them result in complicated Crk localization variations. The Crk localization is further complicated by the fact that the majority of Crk molecules localize to cytosol, which complicates PALM imaging. Even with TIR imaging, the high amount of cytosolic background reduces the effective sampling rate of membrane localized Crk, increases the data collection time, and makes it more challenging to estimate drift directly from the single-molecule data. To validate BaSDI algorithm with the experimental Crk PALM dataset, we added fluorescent beads based fiducial markers into the cell samples during the PALM data collection, and experimentally measured the drift of the sample stage during data collection using the positions of the fluorescent beads. The drift is then also estimated with BaSDI using the single molecule coordinates excluding any information from the fluorescent beads. The comparison between the experimentally measured drift and the BaSDI estimated drift (Fig 2.5) showed remarkable consistency between the two

.

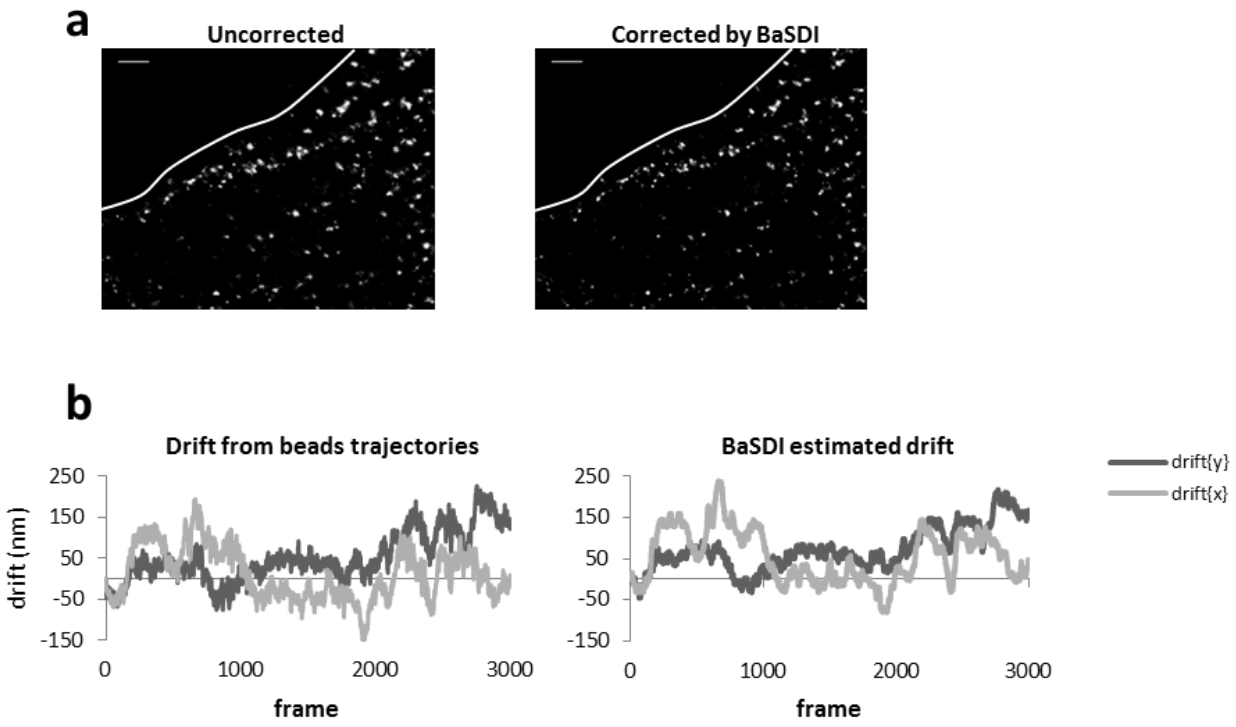

Fig 2.5. BaSDI gives comparable drift estimates to those estimated by fluorescent beads in real PALM experiments. (a) PALM images of mEos3-Crk in an MCF-7 cell before (left) and after (right) drift correction by BaSDI. Scale bars represent 2μm (b) Drift traces estimated by either fluorescent beads (left) or BaSDI (right) in the y (vertical) and the x (horizontal) directions.

with only an estimated RMS error of 25 nm difference between them, and thus we conclude that BaSDI is able to correctly estimate drift from experimental PALM dataset.

2.4.3 Performance of BaSDI algorithm.

Next we tested whether BaSDI achieves better accuracy over the existing image-correlation methods. For comparisons, we generated 100 sets of simulated drift traces and the accompanying PALM datasets each including 1000 individual images. We then estimated the drift using either BaSDI or the image correlation method (Fig 2.6-2.7) using the same ground truth image generated in Fig1 of a size 500x500 pixels. The accuracy of either method was evaluated by computing the error of the estimated drift (by comparing to the input drift) in the mean square error (m.s.e.) sense. Furthermore, to account for different image conditions, we generated simulated data in two different scenarios: a low-sampling-rate scenario in which an average of 50 molecules were detected in each frame, and a high-sampling-rate scenario where 200 molecules per frame were detected. Because image correlation only directly computes drift for a subset of "key" frames, we computationally generated the rest of the drift traces using linear interpolation (Fig 2.6a and 2.7a, right). Another factor that we needed to consider when carrying out the image-correlation was deciding how many substacks the whole dataset should be divided into. Using too many substacks will cause high error in drift estimation between substacks, while too few substacks will lead to fewer data points and a worse temporal resolution(Geisler et al. 2012). For real experimental data, the optimal number can be difficult to determine objectively. For our simulation, we computed the error for 4 different image-correlation groupings (50, 100, 200, and 250 frames). The optimal grouping is empirically determined and the result from it is compared with the error from BaSDI, for both the low-sampling-rate (Fig 2.6) and the high-sampling-rate scenarios (Fig 2.7). As expected, the high-sampling-rate is correlated with lower error for both

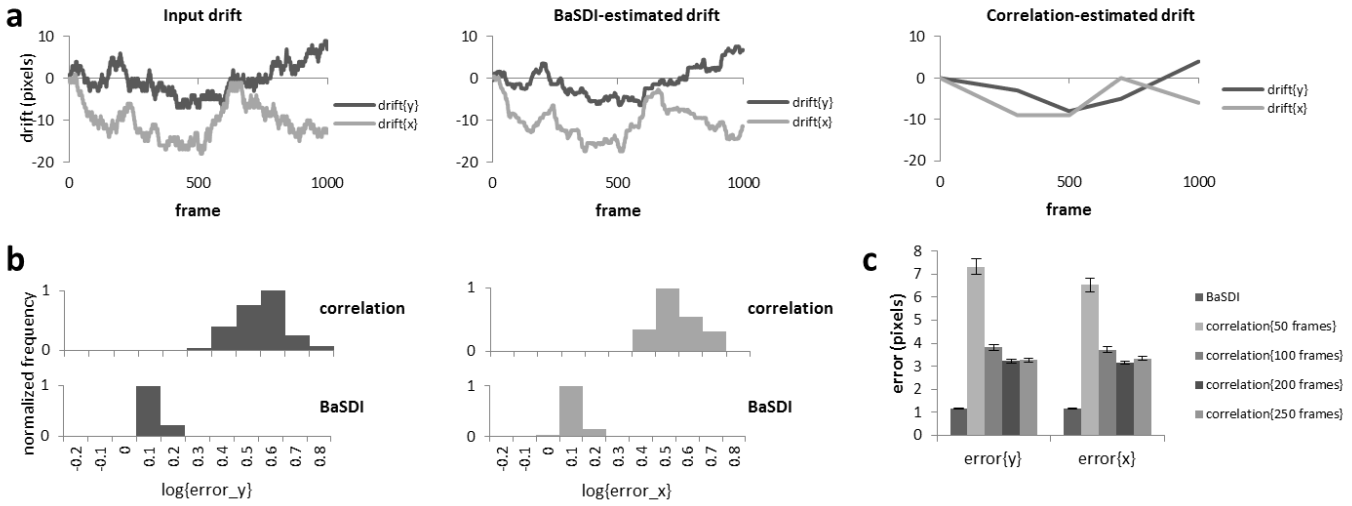

Fig 2.6. Comparing drift estimations by BaSDI and image-correlation at low-sampling-rate (50 molecules/frame). (a) Simulated input drift (left) and drift traces estimated by either BaSDI (middle) or image-correlation (right) using the optimal frame grouping with the least error for the image-correlation in the y (vertical) and the x (horizontal) directions. (b) Histograms comparing the root mean square error distributions of BaSDI and image-correlation in the y and the x directions. The distributions of error were computed from 100 independent runs with randomly generated drifts. (c) Comparing the averaged errors of BaSDI and image-correlation given the different frame groupings. Error bars are s.e.m, n = 100.

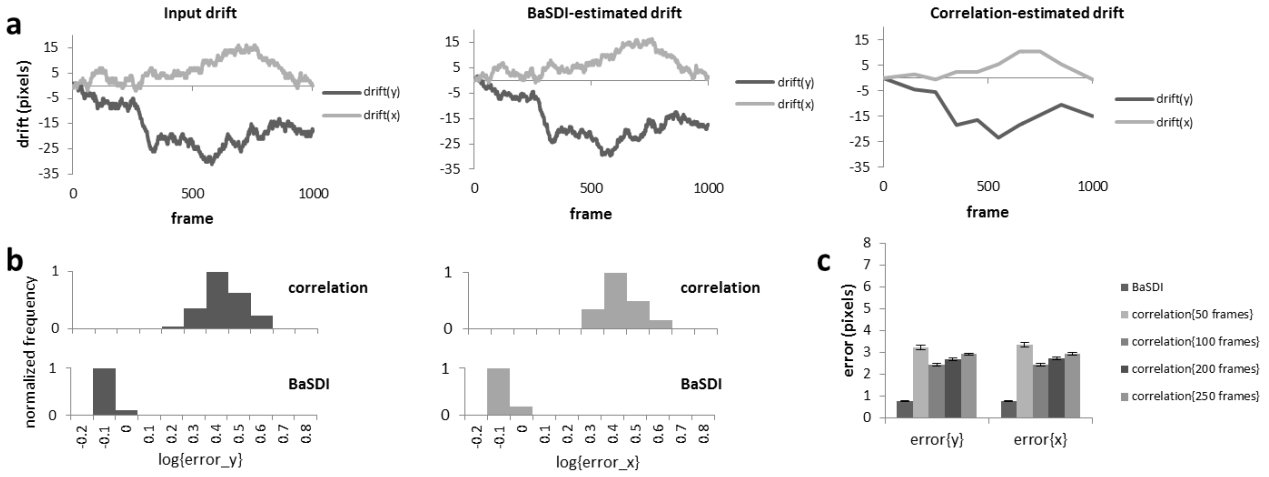

Fig 2.7. Comparing drift estimations by BaSDI and image-correlation at high-sampling-rate (200 molecules/frame). (a) Simulated input drift (left) and drift traces estimated by either BaSDI (middle) or image-correlation (right) using the optimal frame grouping with the least error for the image-correlation in the y (vertical) and the x (horizontal) directions. (b) Histograms comparing the root mean square error distributions of BaSDI and image-correlation in the y and the x directions. The distributions of error were computed from 100 independent runs with randomly generated drifts. (c) Comparing the averaged errors of BaSDI and image-correlation given the different frame groupings. Error bars are s.e.m, n = 100.

methods (Fig 2.6c and 2.7c). Nevertheless, in both cases, we found that BaSDI outperforms image correlation by about 3-4 folds (Fig 2.6b and 2.7b). Interestingly, BaSDI also seems to have a more stable performance, as indicated by the narrow distribution of errors in the histograms (Fig 2.6b and 2.7b) in comparison to the image correlation method.

Finally, we evaluated the relative performance of BaSDI and image-correlation by comparing the image quality from real PALM experimental data (Fig 2.8). To do that we imaged mEos3-Crk in propagating mouse embryonic fibroblast (MEF) cells. In this condition, Crk localizes to the classical strip pattern typical of FA structure. We then applied either BaSDI or image-correlation to carry out the drift compensation (Fig 2.8) based on estimating drift traces of 18000 points. Both methods improved the PALM image sharpness in comparison to the uncompensated result (Fig 2.8). To choose a hyperparameter σ^2, we utilized the existing fluorescent beads tracking data measured separately on the same microscope, and calculated the variance of position drifts between frames. We note that for small σ^2 values, the accuracy of the drift calculation is insensitive to its exact value (Fig 2.9). Therefore, the variance measurement does not need to be very accurate, and the day-to-day variations in the drift characteristics of the microscope are expected to have little impact on the inference results. Thus we did not perform recalibration for the PALM data acquired on different dates. Comparison between the estimated drift from BaSDI and from the correlation method showed that the two methods generally agree with each other except that BaSDI generated not only estimation on key frames but also detailed traces in between key frames, resulting in better resolved image features in the final constructed PALM images (Fig 2.8c).

In all the testing cases, we found that the convergence of the BaSDI algorithm was typically reached in 5 - 10 iterations. For estimating 1000 drift frames from a dataset of 1000 × 1000

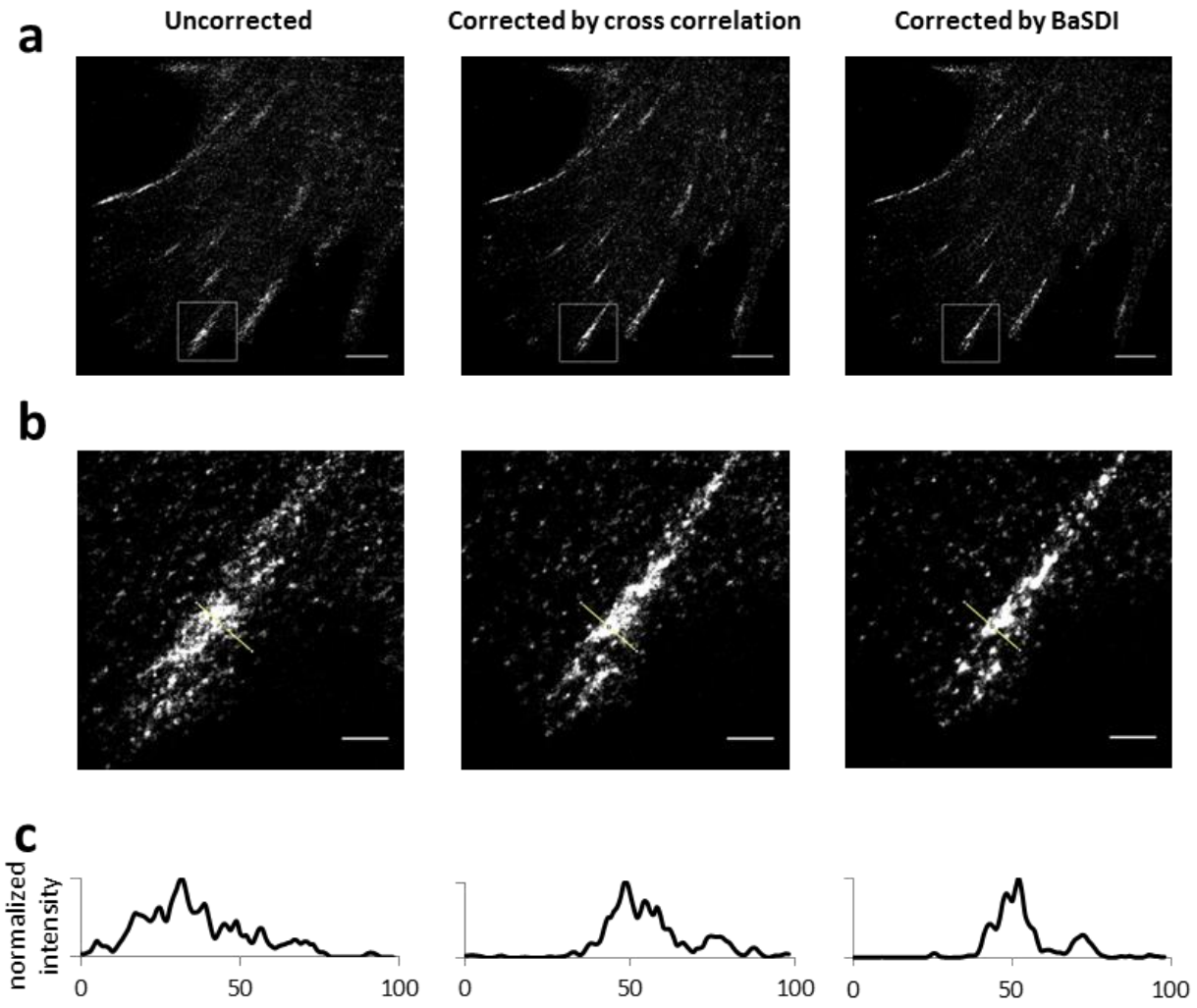

Fig 2.8 . Comparing BaSDI and image-correlation in compensating drift from experimental PALM data. (a) Uncorrected (left), image-correlation corrected (middle) and BaSDI corrected (right) PALM images of mEos3-Crk in a mouse embryonic fibroblast cell. Data showed aggregated localization patterns consistent with focal adhesions. Scale bars represent 5μm (b) Zoomed-in view of the corresponding PALM images at the top, giving detailed view of the focal adhesion structure outlined by the white box in (a). Scale bars represent 1μm. (c) Intensity line-scans from (b) showing the improved image sharpness of the PALM images after drift corrections.

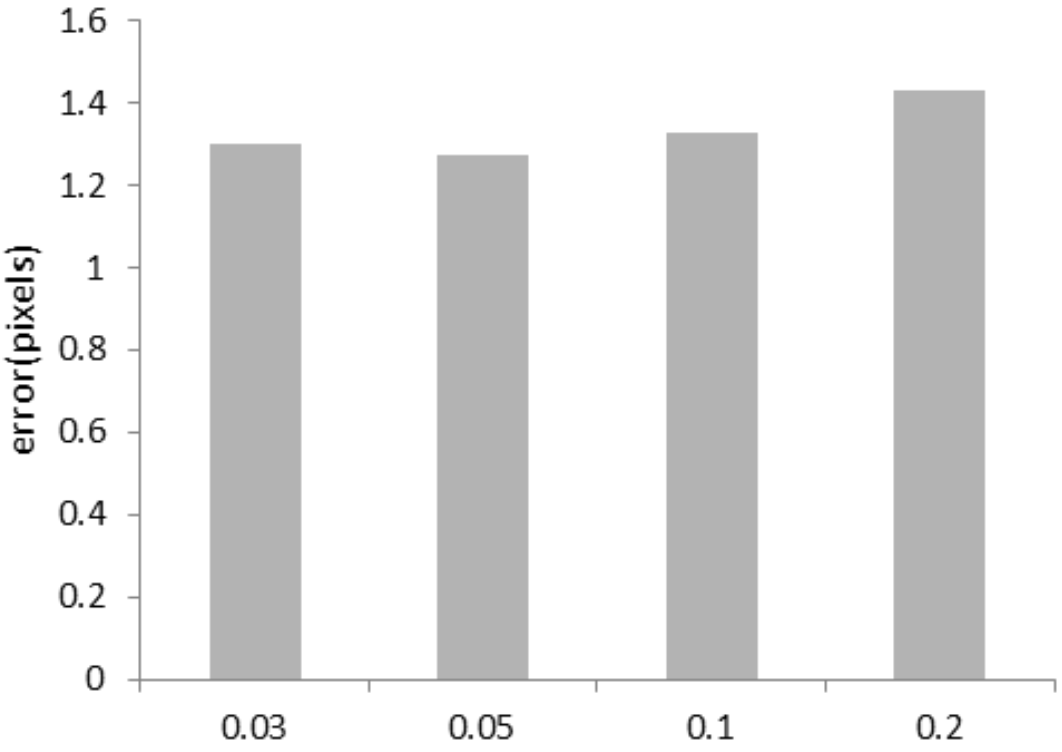

Fig 2.9. BaSDI accuracy is insensitive to the prior parameter. Input drifts were generated with a constant $\sigma^2 = 0.1$. The inferences were computed at σ^2 varying between 0.03 and 0.2. The average errors in the drift interference are relatively constant.

pixels, each iteration takes less than 30 sec to finish on a laptop personal computer. The computational time scales with respect to the number of image frames as O(n), and image size as $O(n^2)$ (Fig 2.10), indicate good scalability to different sizes of datasets.

In conclusion, we have devised an algorithm to estimate sample drift from single-molecule based super-resolution imaging methods. The algorithm is validated by both computer simulations as well as experimental data. The algorithm outperforms previous analysis methods, and more importantly, is based on a statistical principle, not heuristics. We believe it should become a useful addition to the toolset of researchers working on super-resolution imaging.

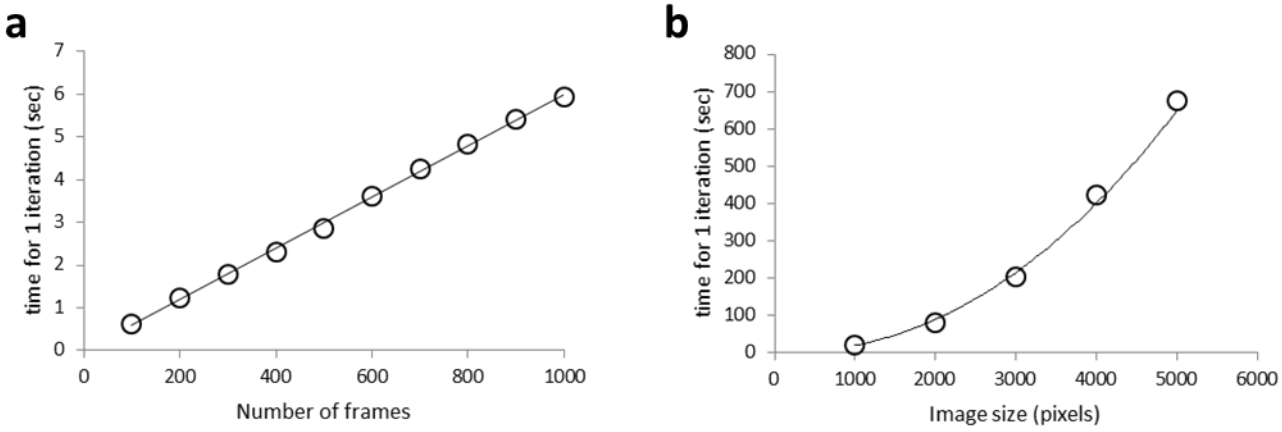

Fig 2.10. Scaling of computational time. (a) Scaling of computation time versus dataset length. (b) Scaling of computation time versus image size.

Chapter 3

Analysis of Measurement Noise in Super-resolution Imaging with Gibbs Sampling

Ahmed Elmokadem and Ji Yu

Author contributions

AE designed and performed the experiments and analyzed data. JY designed the project and performed experiments. AE and JY contributed equally to the writing of this chapter

3.1 Abstract

Single-molecule localization based super-resolution microscopy techniques have revolutionized the optical imaging field. However, like all experimental methods, single-molecule localization microscopy is a technique that comes with measurement noise. Understanding the amount of noise in a given image obtained by single-molecule localization microscopy is a challenging problem. There are two sources of noise: First, there are uncertainties in the exact locations of each molecule detected. Second, multiple detections of the same molecule produce counting errors. The combination of the two typically results in apparent blurred cluster features in the final rendered images. Here we analyze theoretically the noise in single-molecule localization images using a Gibbs sampling approach. The analyses lead to two conclusions that were not previously appreciated: (1) To achieve super-resolution with high statistical confidence, one needs a significantly higher labeling density than what was typically reported in literature. (2) Current image rendering algorithms unnecessarily deteriorate image sharpness (and thus, apparent resolution), especially at high labeling density. Instead we recommend a rendering approach based on statistical sampling that does not have this problem.

3.2 Introduction

Since the introduction of single-molecule localization microscopy (SMLM) techniques over a decade ago (Eric Betzig et al. 2006; Rust, Bates, and Zhuang 2006; Hess, Girirajan, and Mason 2006; Sharonov and Hochstrasser 2006), they became integral to the field of cell biology. One of the key differences between SMLM and other super-resolution fluorescence microscopy techniques like stimulation emission depletion (STED) microscopy (Klar and Hell 1999) and saturated structured illumination microscopy (SSIM) (Gustafsson 2005) is that these latter methods depend on observing subcellular molecular ensembles, while SMLM is able to reach single-molecule localization level.

The driving force behind SMLM is the fact that localizing the centroids of single molecules can be achieved with a precision that is much higher than the ~200 nm Abbe resolution limit of regular optical microscopy (Huang, Bates, and Zhuang 2010; Yildiz et al. 2003). It is crucial then to dilute the fluorescence signal in a biological sample to the point that single molecules are discernible and SMLM methods do that by extending the acquisition time and allowing for only a subset of molecules to be detected at a given frame either through the use of photoactivatable fluorescent proteins or dyes (PALM/FPALM/STORM) (Eric Betzig et al. 2006; Hess, Girirajan, and Mason 2006; Rust, Bates, and Zhuang 2006) or diffuse fluorescent probes (PAINT) (Sharonov and Hochstrasser 2006). In the final stage, these single-molecule localization data are stacked together to construct the final super-resolution image.

SMLM methods still face several challenges, among which are the poor temporal resolution and sample drift (Elmokadem and Yu 2015). Another challenge that is the focus of this chapter is the measurement noise. One source of the noise is the localization uncertainty. According to Eq.

1.1, this uncertainty is dependent on the number of photons collected with more photons meaning less uncertainty about the fluorophore's location. In addition, because of the extended acquisition of SMLM, a single molecule could be detected multiple times either due to the phenomenon of photoblinking in case of PALM/FPALM/STORM or because of the transient binding of diffuse probes in case of PAINT. As a result, every time the molecule is detected, the uncertainty in its localization would lead to assignment of new molecule coordinates that are offset from where it should be, causing blurring of the feature of interest. This also plays a role in the 'apparent' cluster morphology observed with most proteins when they are imaged with SMLM (Garcia-parajo et al. 2014).

Because of the unavoidability of localization uncertainty in SMLM, it is crucial to correctly approach this problem to get more accurate super-resolution images. Unfortunately, all the current SMLM image rendering algorithms deal with multiple detections of the same fluorophore as separate events and would simply fit every localization coordinates with a 2d Gaussian spot whose width is governed by the localization uncertainty. This approach fails to capture the reality that each one of these detections could be in fact coming from the same fluorophore or the same location.

In addition to localization uncertainty, another factor that greatly influences the quality of the final super-resolution image is the labeling density, which is the number of fluorescent labels per unit of space of the sample (Shroff et al. 2008). A high enough labeling density is crucial to fully capture the structure of the feature of interest as insufficient labeling might cause a continuous structure to appear discontinuous.

Previous attempts have been made to find a mathematical description of the effect of labeling

density on resolution and one popular approach is using the Nyquist criterion, which states that in order to reach a certain resolution d, we need to have a spatial labeling frequency f that is half of that desired resolution; i.e.: $f = d/2$ (Shroff et al. 2008). Unfortunately, the heuristic nature of this approach highlights the need for a more rigorous mathematical description. Also, it doesn't consider the heterogeneity of biological structures that might lead to disparate local densities.

Here we present a novel statistical approach to analyze the measurement noise in images rendered by SMLM. Our approach employs Gibbs sampling to sample from all possible molecular localizations based on a certain probability distribution that considers the fact that subsets of close localizations might be arising from the same fluorophore. The generated samples allow us to extract useful statistics such as how confident we are that two features in the super-resolution image are really resolved by incorporating the information from both localization and labeling density uncertainties. Our algorithm gave us a more rigorous quantification of the influence of these uncertainties on the resolving power and the averaged image of all samples showed a significantly improved resolution than that obtained by the regular image rendering algorithm. Furthermore, our algorithm was able to estimate uncertainties in cluster size measurements of clustering proteins as well as differentiate between real and apparent clusters.

3.3 Methods

3.3.1 Statistical model

Let $\boldsymbol{\theta}^{\circ} \in \boldsymbol{R}^{mn}$ represent the super-resolution image of the size $m \times n$, so that θ_i° is the concentration of the target molecules at the i-th pixel. Here we assume that the localization is

performed in two dimensions, but extension to three-dimensional localization should be straightforward. Furthermore, of particular practical interest is its normalized form: $\boldsymbol{\theta} = \boldsymbol{\theta}^{\circ}/\Sigma_{\mathrm{i}}\theta_i^{\circ}$. Also let $\boldsymbol{C} \in \boldsymbol{N}^K$ represent the pixels indices (as in $\boldsymbol{\theta}$) of all K number of single molecules detected and localized. Therefore, $\boldsymbol{C}$ is the experimental data, and $\boldsymbol{\theta}$ is the image we are trying to create from the data. As the single-molecule localizations are not without noise, the molecular distribution $\boldsymbol{\theta}$ is linked to the experimental data $\boldsymbol{C}$ only in a probabilistic sense. To obtain an analytical form of the distribution function, we introduce an experimentally unobserved hidden variable $\boldsymbol{D} \in \boldsymbol{N}^K$**,** which represents the *true* locations of the K detected single molecules. Thus:

$$\begin{aligned} P(\boldsymbol{\theta}|\boldsymbol{C}) &= \frac{1}{Z}P(\boldsymbol{C}|\boldsymbol{\theta}) = \frac{1}{Z}\sum\nolimits_{\boldsymbol{D}} P(\boldsymbol{C}|\boldsymbol{D})P(\boldsymbol{D}|\boldsymbol{\theta}) \\ &= \frac{1}{Z}\sum\nolimits_{\boldsymbol{D}} \prod\nolimits_k \mathcal{N}\left(\mathfrak{D}_{C_k,D_k}; \sigma_k\right)\theta_{D_k} \end{aligned} \tag{3.1}$$

where $\mathcal{N}(\,)$ is the normal distribution, $\mathfrak{D}_{i,j}$ is the distance between pixels i and j, σ_k is the single molecule localization uncertainty, and Z is a normalization constant.

In principle, the probability distribution in Eq. 3.1 encompasses the complete statistical information allowing computation of any statistical quantities of $\boldsymbol{\theta}$, including, most notably, its first moment, which is also the expectation value for $\boldsymbol{\theta}$, $\boldsymbol{E}_{\boldsymbol{\theta}} = \int \boldsymbol{\theta} \cdot P(\boldsymbol{\theta}|\boldsymbol{C})d\boldsymbol{\theta}$, or its second moment, which is related to the uncertainty (noise) of the computed $\boldsymbol{\theta}$ expectation values. Unfortunately, exact analytical calculation of the probability distribution in Eq. 3.1 is essentially impossible due to the high-dimensional nature of the problem. Instead we propose performing approximate numerical calculations by drawing samples of $\boldsymbol{\theta}$ from the distribution as means to approximate the complete distribution. After we draw a sufficiently large number of samples from the probability distribution in Eq. 3.1: $\boldsymbol{\theta}^1, \boldsymbol{\theta}^2, \ldots, \boldsymbol{\theta}^n$, numeric estimation of the various

moments of the random variable can be computed with relative ease. For example, computing the expectation value for $\boldsymbol{\theta}$ can be done as follows:

$$\mathbb{E}[\boldsymbol{\theta}] \approx \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{\theta}^i \tag{3.2}$$

3.3.2 Gibbs sampling

A class of sampling methods that revolutionized the field of statistics is known as the Markov chain Monte Carlo (MCMC) and an important member of this class is Gibbs sampling (Glickman and van Dyk 2007; Hazan et al. 2016; Geman and Geman 1984). The power of Gibbs sampling comes from the fact that it breaks down a high-dimensional sampling problem into several one-dimensional sampling problems that are much easier to approach. Suppose we have a parameter vector $x = (x_1, x_2, \dots, x_k)$ that we want to sample from. Gibbs sampling provides a mechanism to sample from this parameter vector by going through cycles where a single element is sampled in each step conditioned on the current values of all other elements instead of sampling directly from the whole high-dimensional vector. This is carried out for enough iterations to reach convergence (Algorithm 3.1).

We make use of this idea to alternatively sample from the high dimensional D and $\boldsymbol{\theta}$ parameters introduced in Eq. 3.1. This is done by first initializing the algorithm with an initial guess for $\boldsymbol{\theta}$ **(**$\boldsymbol{\theta}^\circ$**).** In our case we generated $\boldsymbol{\theta}^\circ$ by simply stacking all localization events together to construct a 2 dimensional histogram. Then the sampling process is carried out in two sequential steps: the first step is to sample from the distribution of $\boldsymbol{D}$ conditioned on a known $\boldsymbol{\theta}$

Algorithm 3.1. Gibbs Sampling

Initialize $\boldsymbol{x}^0$

for iteration $i = 1,2, \dots, n$ **do**

$x_1^i \sim P\left(x_1 \middle| x_2^{i-1}, x_3^{i-1}, \dots, x_k^{i-1}\right)$

$x_2^i \sim P\left(x_2 \middle| x_1^i, x_3^{i-1} \dots, x_k^{i-1}\right)$

.

.

.

$x_k^i \sim P\left(x_n \middle| x_1^i, x_2^i, \dots, x_{k-1}^i\right)$

end for

followed by sampling from the distribution of $\boldsymbol{\theta}$ conditioned on $\boldsymbol{D}$. The first distribution can be written as follows:

$$P(\boldsymbol{D}|\boldsymbol{\theta},\boldsymbol{C}) = \prod_k \frac{\mathcal{N}\left(\mathfrak{D}_{C_k,D_k};\sigma_k\right)\theta_{D_k}}{\sum_j \mathcal{N}\left(\mathfrak{D}_{C_k,j};\sigma_k\right)\theta_j} \tag{3.3}$$

We notice that Eq. 3.3 has a completely separable form that allows for the application of Gibbs sampling.

The second distribution is written in the Bayesian form as:

$$\begin{aligned} P(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{C}) &= \frac{P(\boldsymbol{\theta})P(\boldsymbol{D},\boldsymbol{C}|\boldsymbol{\theta})}{P(\boldsymbol{D},\boldsymbol{C})} \\ &= \frac{P(\boldsymbol{\theta})P(\boldsymbol{D}|\boldsymbol{\theta})P(\boldsymbol{C}|\boldsymbol{D},\boldsymbol{\theta})}{P(\boldsymbol{D},\boldsymbol{C})} \\ &= \frac{P(\boldsymbol{\theta})P(\boldsymbol{D}|\boldsymbol{\theta})P(\boldsymbol{C}|\boldsymbol{D})}{P(\boldsymbol{D},\boldsymbol{C})} \end{aligned} \tag{3.4}$$

We can drop the constant terms with respect to $\boldsymbol{\theta}$ to get:

$$\begin{aligned} P(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{C}) &\approx P(\boldsymbol{\theta})P(\boldsymbol{D}|\boldsymbol{\theta}) \\ &\approx P(\boldsymbol{\theta})\prod_k \theta_{D_k} \end{aligned} \tag{3.5}$$

where $P(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{C})$ is the posterior distribution of $\boldsymbol{\theta}$, while $P(\boldsymbol{\theta})$ is the prior distribution.

3.3.3 Prior distribution of $\boldsymbol{\theta}$

The proper choice of the prior distribution $P(\boldsymbol{\theta})$ in Eq. 3.5 could be an important factor in simplifying the computation of the posterior distribution $P(\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{C})$. We use an uninformative Dirichlet distribution for the prior parametrized with an $\boldsymbol{\alpha}$ $\boldsymbol{1}$vector: $\boldsymbol{\theta} \sim Dir(\boldsymbol{\alpha})$. As a result, the

posterior distribution also assumes the form of a Dirichlet distribution with a different parameterization: $\boldsymbol{\theta}|\boldsymbol{D},\boldsymbol{C} \sim Dir(\boldsymbol{\alpha}+\boldsymbol{D})$ and hence, sampling from that distribution can be done with relative ease.

3.3.4 Generation of simulated data

We used Monte Carlo simulation to generate random localization datasets that correspond to the observed data $\boldsymbol{C}$ in a real experimental setup. This is done by using ground-truth images with specific features that govern where the molecules are most likely to be located. Localization uncertainty was introduced via Gaussian noise to the feature locations and then random molecule localizations are sampled accordingly. The extent of localization uncertainty is determined by the standard deviation of the Gaussian noise.

3.3.5 Convergence

Gibbs sampling is guaranteed to converge to the true distribution after a burn-in phase (Asmussen and Glynn 2011). We tested convergence by monitoring the cumulative mean of intensity of randomly picked pixels as more samples are being acquired. It is assumed that convergence is reached when these cumulative means don't change by much anymore (Fig. 3.1). We notice that convergence is reached faster for background pixels than foreground pixels. For all of our runs, we sampled 50000 samples and in general the burn-in phase lasted for the first 20000 samples of each run.

3.3.6 Cluster size analysis

In the cluster analysis part, we analyzed the cluster sizes for each sample to get the distribution of these sizes. To do so, we used thresholding where a fixed intensity value is used

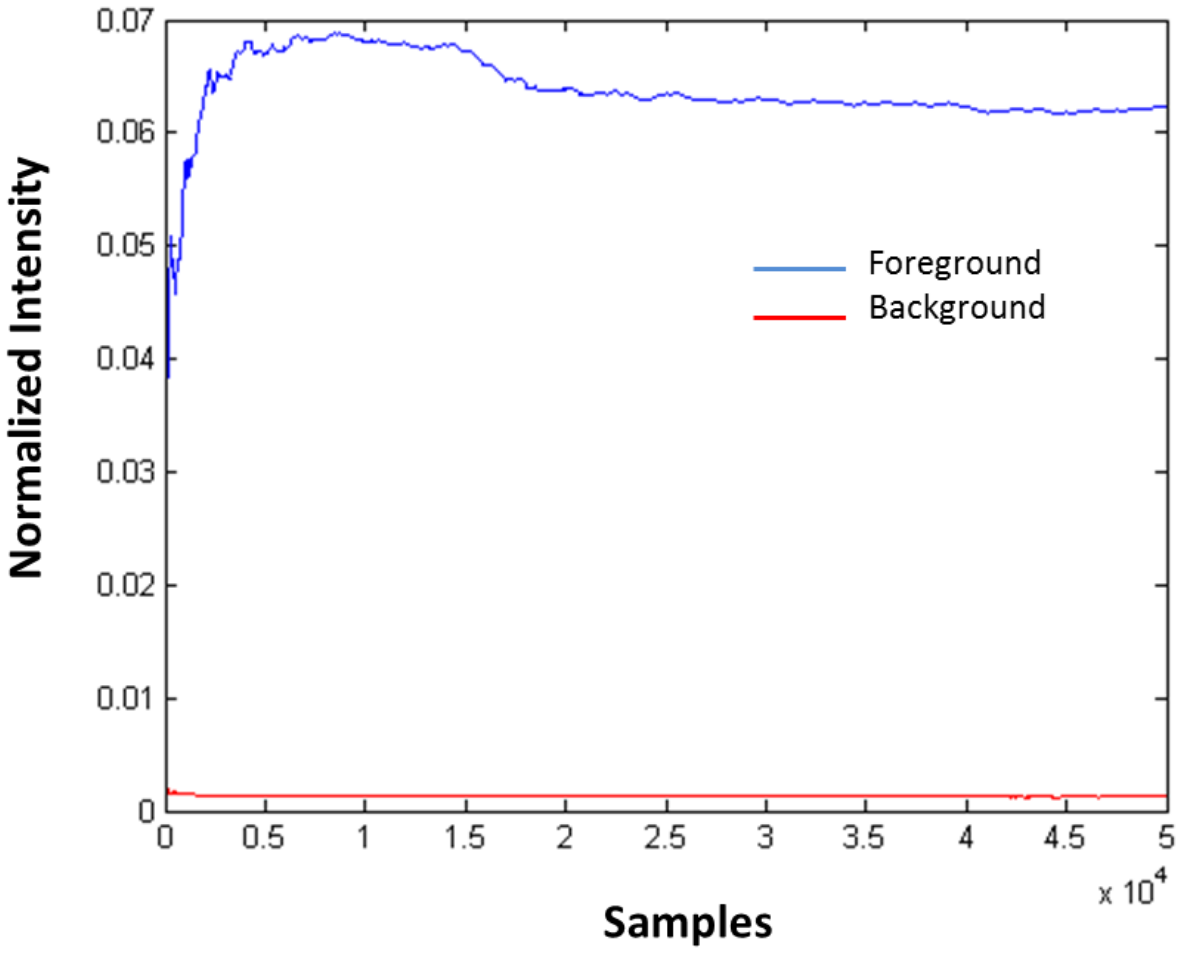

Fig. 3.1. Algorithm convergence. The figure shows cumulative averages of pixel intensities for pixels in the foreground (blue) and the background (red) regions.

to filter out all pixels with lower intensity. After that we count all the remaining pixels and sum them together and this is the area of the cluster a. By assuming that a cluster is a circle in 2D, we can use the formula: $d = 2\sqrt{\frac{a}{\pi}}$ to compute the cluster diameter. In general the thresholding value was 13.5% ($1/e^2$), which is the value commonly used in calculating the beam width of a laser. Nevertheless, we used visual inspection to adjust the thresholding value when 13.5% isn't satisfactory as with the larger cluster sizes where this value would lead to filtering out a lot of the pixels constituting the real cluster size, which in turn leads to underestimation.

3.3.7 Improving algorithm efficiency

As all sampling methods, Gibbs sampling suffers from the drawback of being slow. To improve the algorithm efficiency we started by analyzing the effect of image size and the number of molecules detected on the computation time (Fig. 3.2). The image size didn't have much effect on the computation time as analyzing a 100 times bigger image only doubles the computation time (Fig. 3.2a).

On the other hand, the computation time increases linearly with the number of molecules, which means that doubling the number of molecules doubles the computation time (Fig. 3.2b). This could lead to a very slow convergence since biological samples with areas of the order of tens of micrometers typically contain $10^7 – 10^8$ molecules.

To circumvent this problem and improve the algorithm efficiency, we notice that both of the conditional distributions in Eq. 3.3 and Eq. 3.5 are in the form of factorized destitutions, which allows for the parallelization of the algorithm. The parallelization was done using the

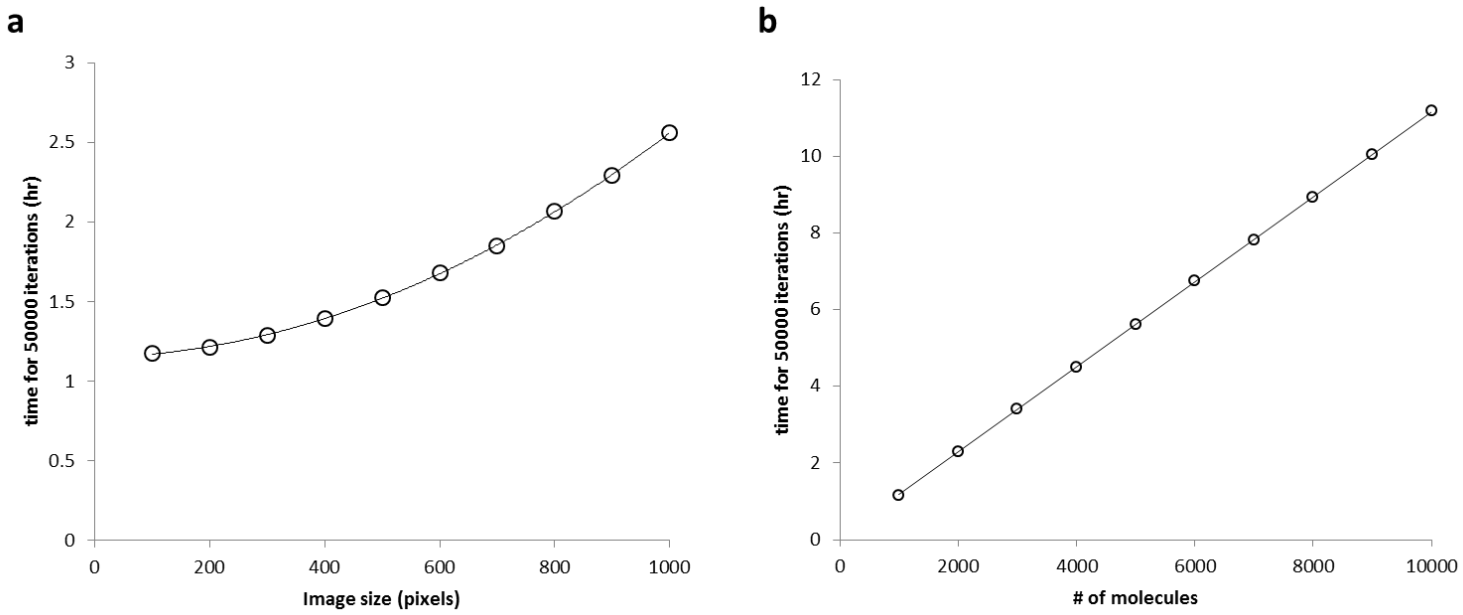

Fig. 3.2. Scaling of computation time. (a) Scaling of computation time vs image size with the quadratic fit (solid line). Number of molecules was kept constant at 1000 molecules. (b) Scaling of computation time vs number of molecules detected with the linear fit (solid line). Image size was kept constant at 100x100 pixels image.

MATLAB® Parallel Computing Toolbox and all jobs were run on the High Performance Computing facility available at the Center for Cell Analysis and Modeling (CCAM), University of Connecticut Health Center. This optimization led to more than seven-fold improvement in the computation speed.

Another optimization step we did was that we binned together localization events appearing within the same coordinate. This improved the computation time dramatically since it allowed for faster sampling from the distribution in Eq. 3.3. The drawback of doing this is that it forces us to use a fixed localization uncertainty for all localization events even though the algorithm is built to handle different uncertainties for different localizations. This assumption is not very accurate since each detected fluorophore is expected to release a different number of photons and hence, the precision of localization would be different every time, but in practice, this difference is very small such that fixing the localization uncertainty wouldn't affect the results by much. As will be shown in the results section, we confirmed this conclusion by running two parallel scenarios with fixed and varying localization uncertainties and the difference in results was negligible.

3.4 Results

3.4.1 Analyzing uncertainty in resolution

To assess our algorithm performance in exploring the localization uncertainty to resolve features and quantify the uncertainty in resolution, we used simulated data and started by analyzing a very simple scenario; two point sources in a 1-dimensional image. As mentioned in the methods section, we used these point sources as the ground truth or the seeds for simulating

localization data by introducing Gaussian noise with a prespecified standard deviation to each localization event arising from these point sources. Since localization uncertainty and distance between features are closely related parameters affecting resolution, i.e.: doubling the distance between the two points is equivalent to halving the localization error, in all of our simulations we fixed the standard deviation of the localization uncertainty and changed the distance between the two point sources to explore different scenarios. This makes intuitive sense since the localization uncertainty of a particular fluorophore isn't expected to change by much whereas the distance between features is expected to be the dominant variable in defining resolution.

Based on that assumption we explored different scenarios by varying distances between the two point sources (60, 90, 120, 150, and 180 nm) in Fig. 3.3-3.7 while maintaining the standard deviation fixed at 30 nm. In each figure we show the ground truth image used for generating the simulated localization data and the comparison between the regular image rendering algorithm, which assigns a 2D Gaussian spot to each localization event, and the average of all the samples generated by our Gibbs sampling algorithm after getting rid of the burn-in samples. We also explored the effect of labeling density (number of molecules/labels per unit area) on resolution by running the simulations with varying number of molecules (20, 60, 200, 600 and 2000 molecules). We used the number of molecules as proxy for labeling density since we fixed all lengths and all areas for the 1D and the 2D simulations, respectively.

By inspecting panel (b) of Fig. 3.3-3.7, we can see an evolving behavior of our algorithm as the number of molecules increases. The performance of our algorithm in terms of resolving power is worst at low number of molecules and improves with increasing the number of molecules as evident from the averages generated by our algorithm as shown in the figures. This is due to the sparsity of information available for our algorithm in a low molecule number

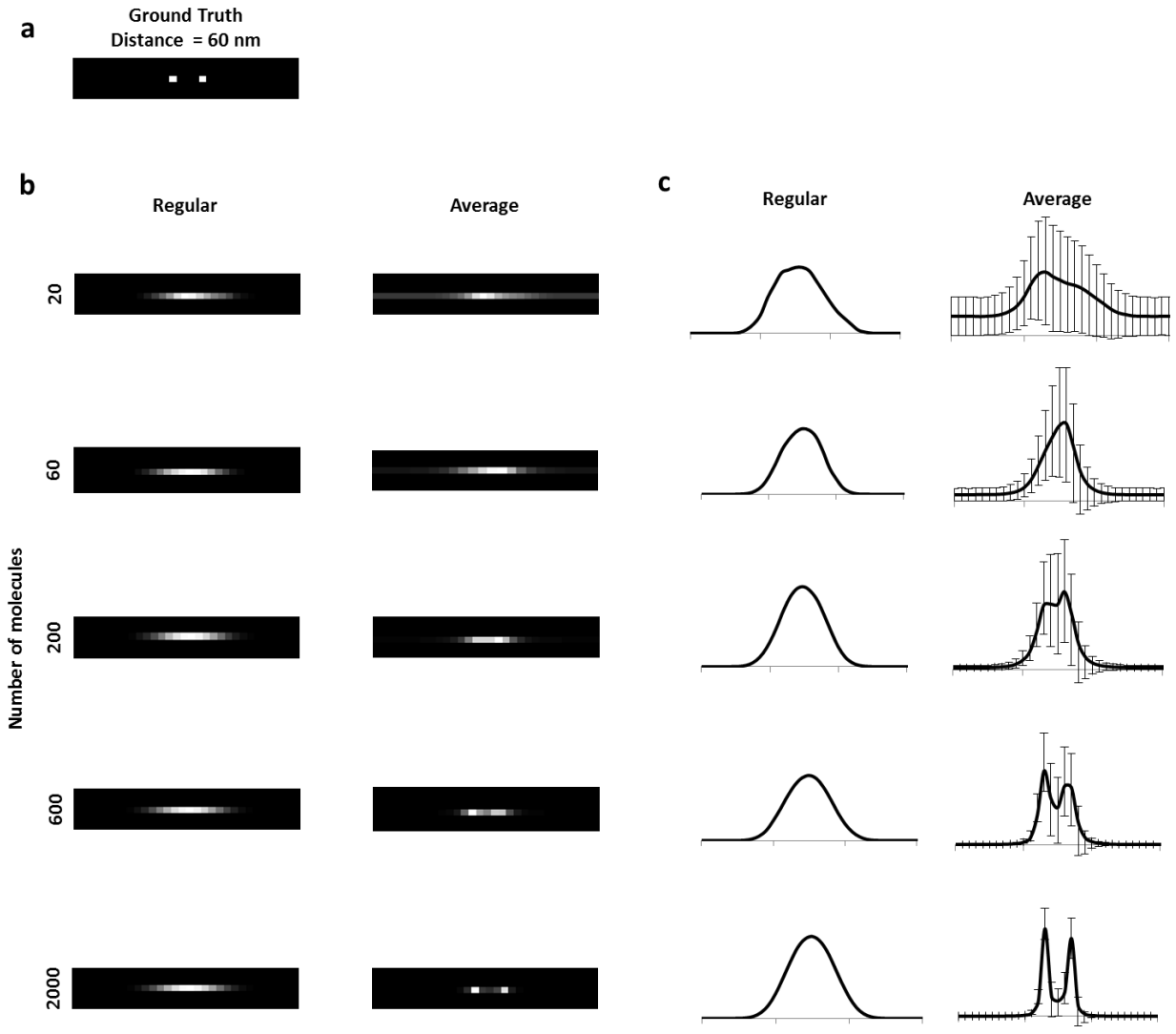

Fig. 3.3. Analysis of two point sources 60 nm apart in a 1D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Length of images are 0.75 um and error bars represent standard deviation.

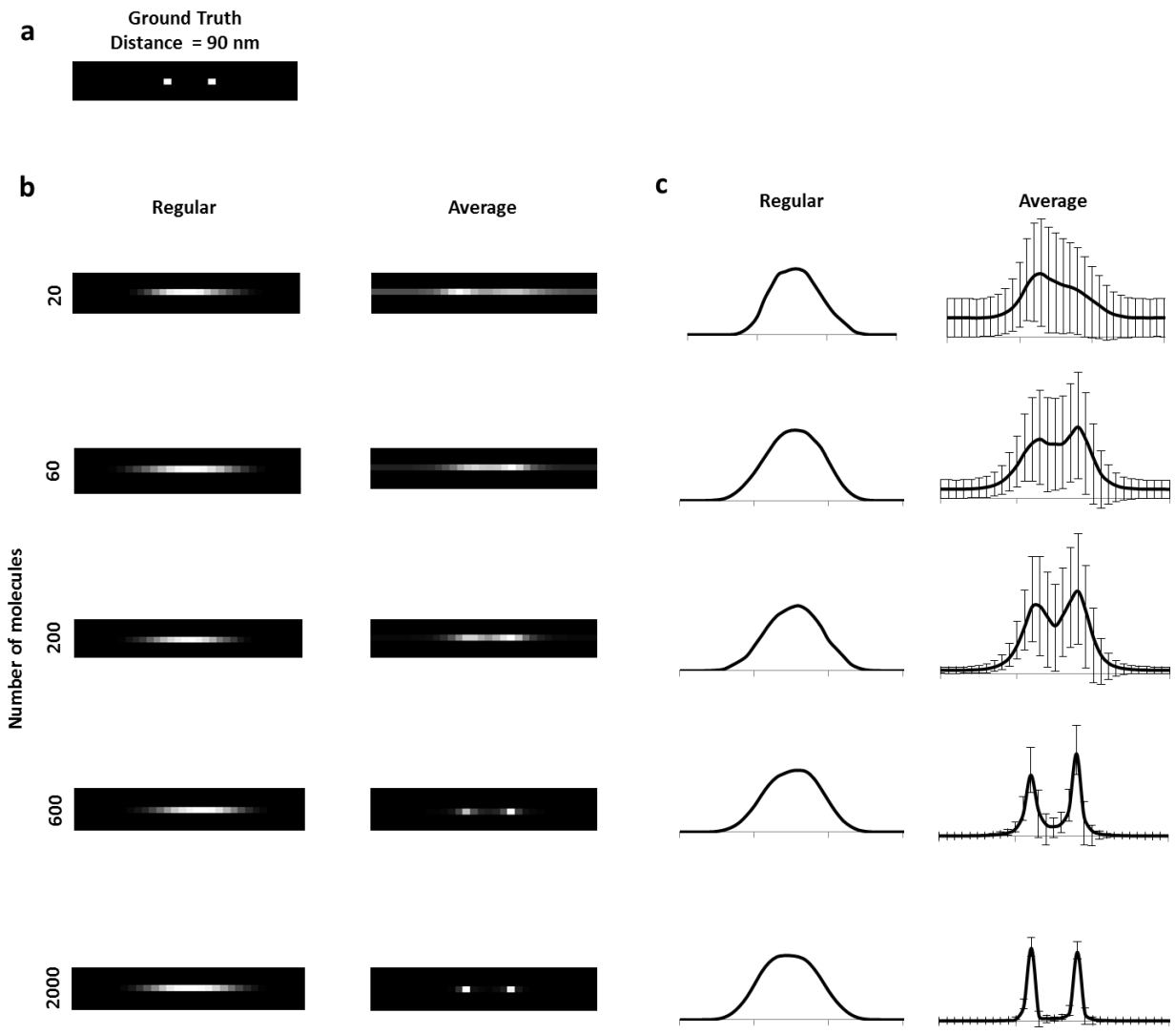

Fig. 3.4. Analysis of two point sources 90 nm apart in a 1D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b).). Length of images are 0.75 um and error bars represent standard deviation.

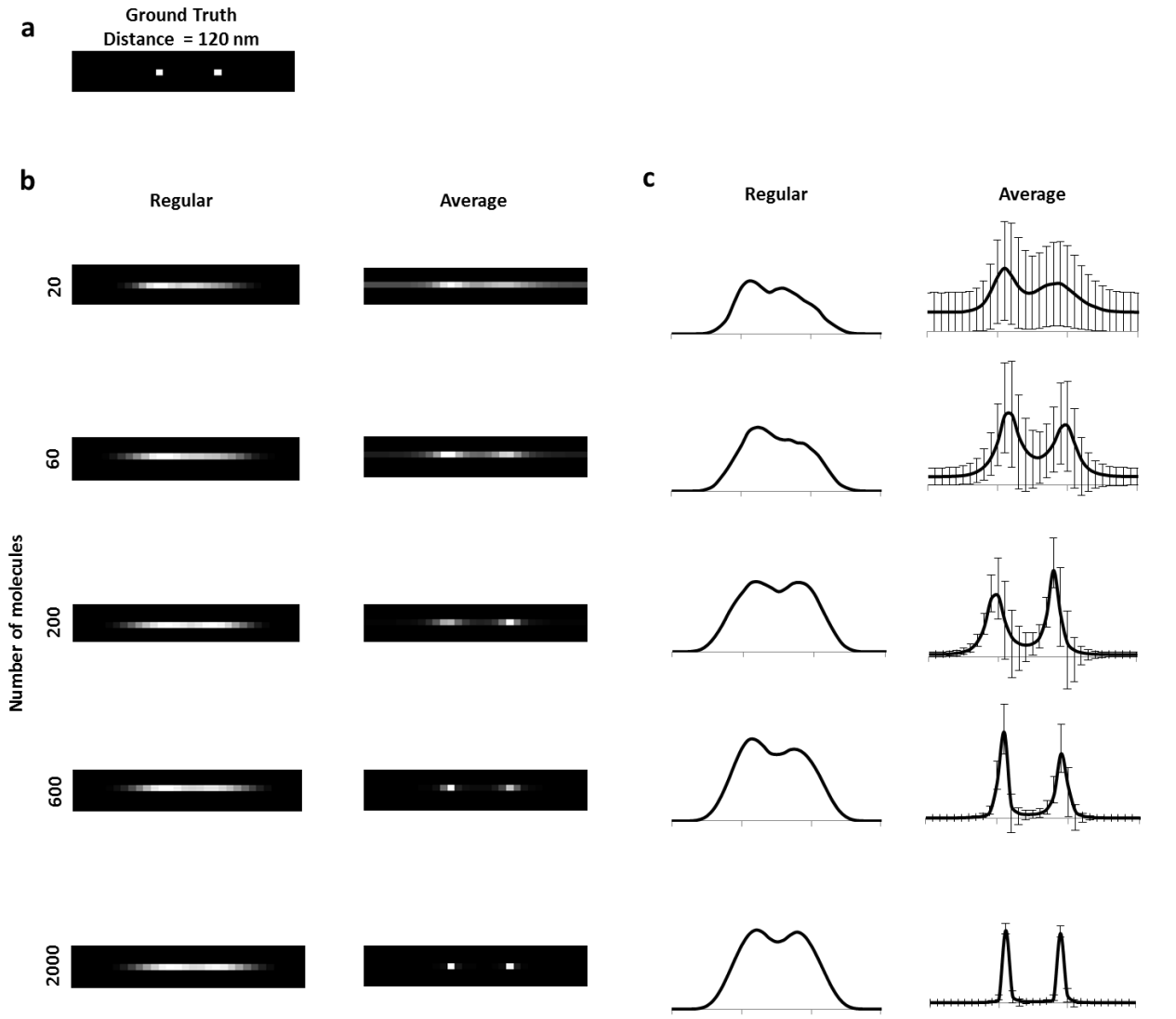

Fig. 3.5. Analysis of two point sources 120 nm apart in a 1D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Length of images are 0.75 um and error bars represent standard deviation.

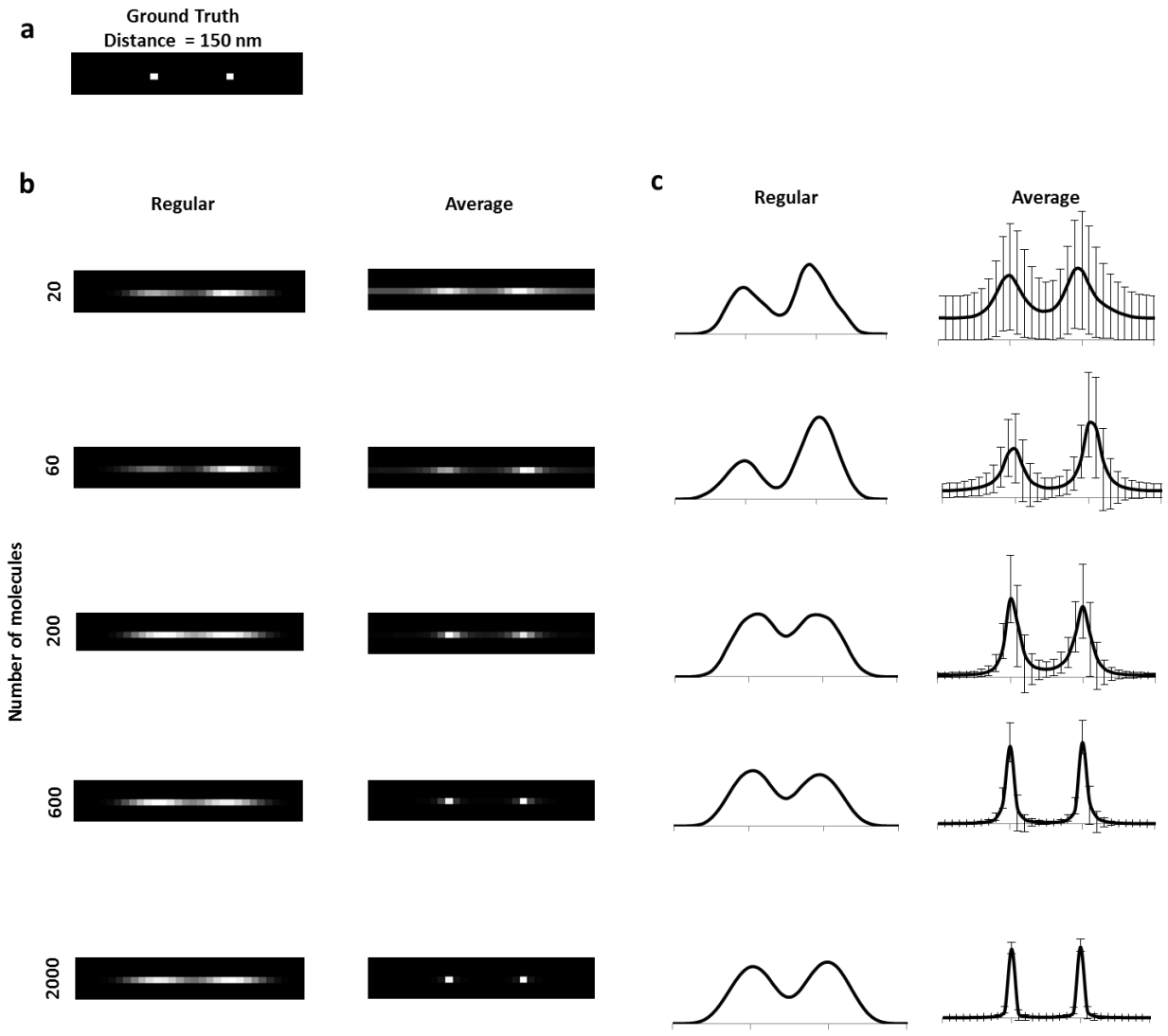

Fig. 3.6. Analysis of two point sources 150 nm apart in a 1D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Length of images are 0.75 um and error bars represent standard deviation.

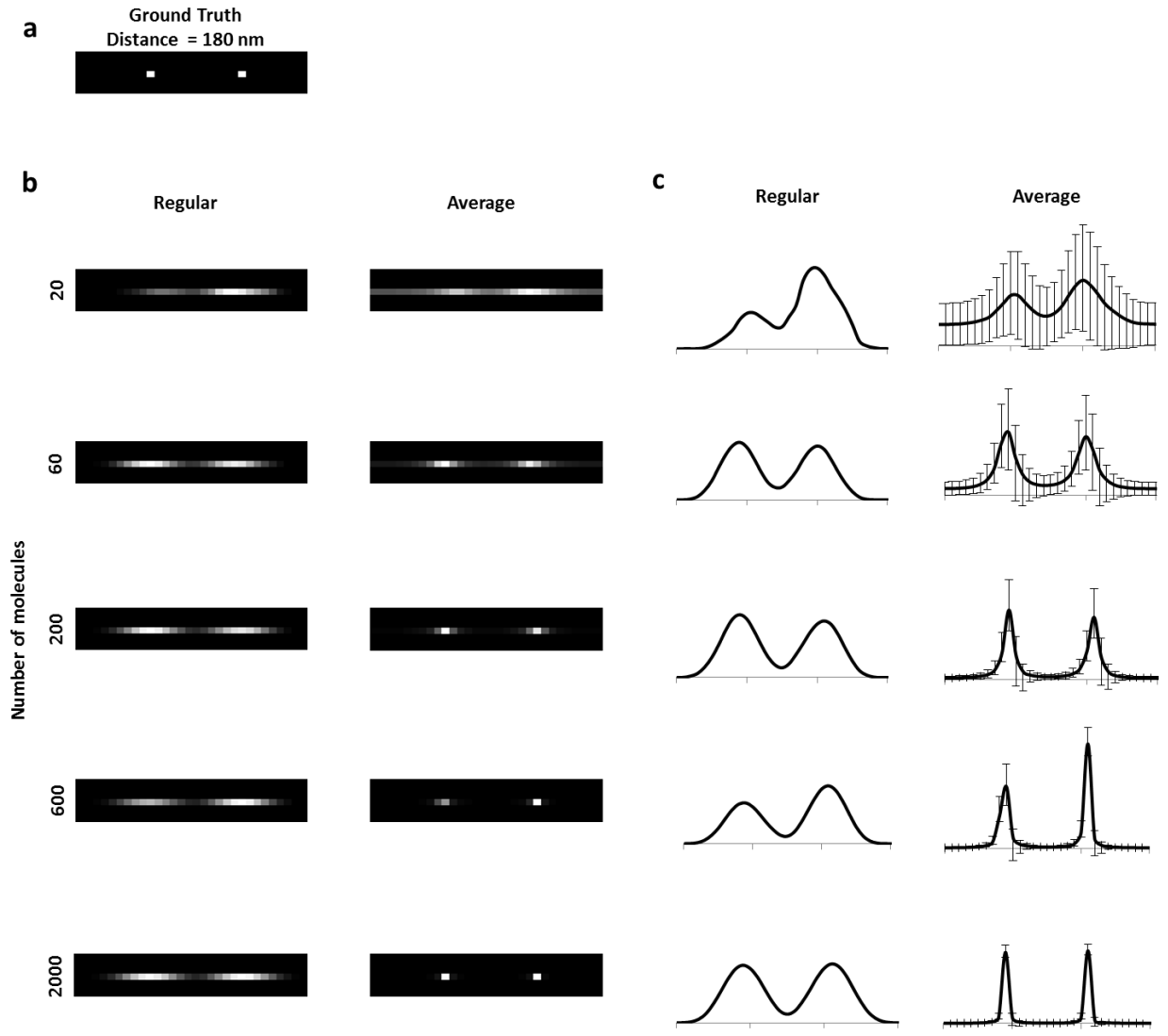

Fig. 3.7. Analysis of two point sources 180 nm apart in a 1D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Length of images are 0.75 um and error bars represent standard deviation.

scenario, which leads to difficulty in inferring the statistical relationships between the subsets of localization events. On the other hand, the more molecules we are detecting, the more information collected and the better resolved images we get. We can clearly see that in Fig. 3.3-3.7 where the features in the averaged images get sharper and become closer to the ground truth images as the number of molecules increases whereas for the regular rendering algorithms, the image doesn't improve with a lot of these images failing to even distinguish between the two features. The same improvement of resolution of the averaged images can also be seen when the distance between the features get larger. It is also worth noting that even at low molecular number where our algorithm is at its worst, it still generates averaged images that are similar to the corresponding images generated by regular rendering algorithms.

The linescans in panel (c) of Fig. 3.3-3.7, which correspond to the images in panel (b), show the unique advantage that our algorithm offers where we can use the generated Gibbs samples to extract useful statistics that provide quantification of uncertainty in the resolution between the two features. We present that by the standard deviation error bars and we can see how the error bars start large at low molecule numbers and shrink as the number of molecules increases. This highlights the importance of labeling density in defining resolution as detecting more molecules dissipate our uncertainty about the underlying feature structure. Of course, regular image rendering algorithms lack this capability and all we can get from the images generated by these are linescans that provide a simple visual guide without any quantification of uncertainty (Fig. 3.3-3.7).

Fig. 3.8 summarizes the effect of the three different parameters: distance between features, standard deviation of localization uncertainty and number of molecules (labeling density) on resolution in the simulated scenarios in Fig. 3.3-3.7. As previously mentioned, the first two

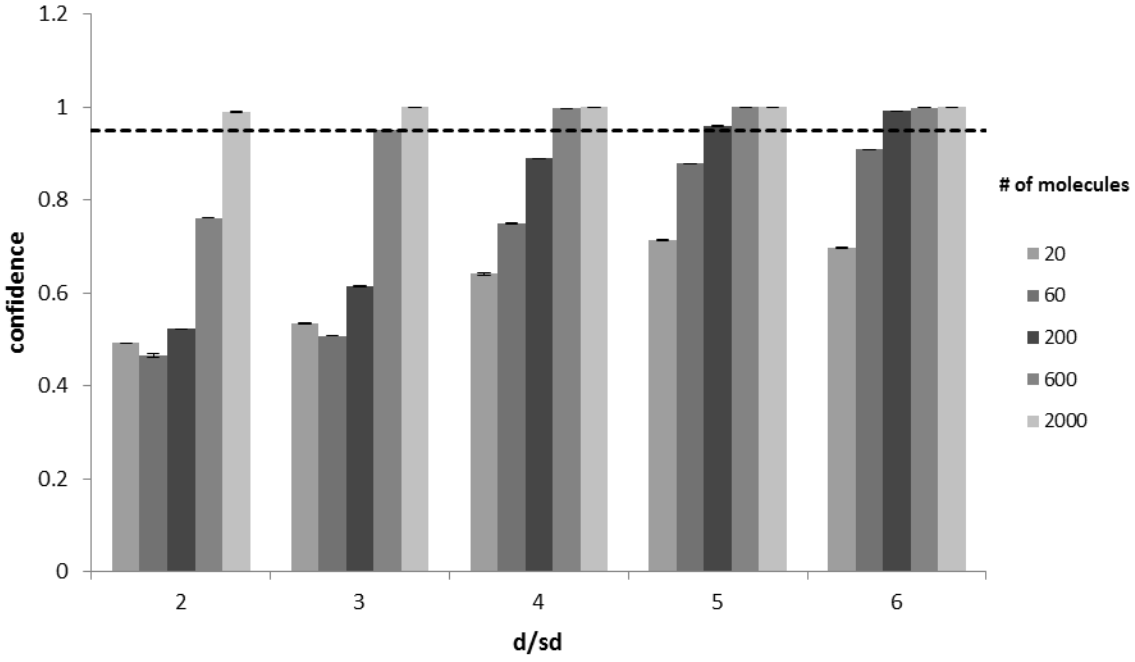

Fig. 3.8. Summary of the 1D simulation results. The bar plots show the confidence in resolving the two features with the increasing number of molecules for each of the *d/sd* (distance to standard deviation ratio) scenarios. Dashed line represents the 95% confidence level and error bars represent the s.e.m, n = 3.

parameters are closely related so we combined them in one parameter named *d/sd* and it represents the ratio between the two parameters. This figure can be used to identify the approximate number of molecules required to reach a 95% confidence level that two features are resolved for a particular 1D scenario. For example, for a *d/sd* = 4, we need a number of molecules that falls between 200 and 600 molecules to reach a 95% confidence level.

We then extended our simulations to two point sources in a 2-dimensional image and we used similar scenarios for distances between features and number of molecules as the 1D simulations (Fig 3.9-3.13). In general, our algorithm's performance was consistent with that of the corresponding 1D scenarios with one main exception, which is it takes more molecules to be confident about the separation between the two features in the 2D scenarios than the corresponding 1D scenarios (Fig. 3.14a). This makes sense since a single pixel in a 1D scenario gets sampled from more often than from a corresponding pixel in a 2D scenario given the same total number of molecules and hence, we are getting more information from the 1D scenario. To get better estimates of the critical number of molecules required to reach resolution significance between the two features, we used linear interpolation between the number of molecules right before and after reaching the 95% confidence mark for every *d/sd* value in Fig. 3.14a.

Interestingly, we saw a pattern that starts at a larger critical number of molecules for small *d/sd* values and then converges to a value of ~300 molecules at larger *d/sd* values (Fig. 3.14b). This means that even at very large separation between features, we still need a relatively large critical number of molecules to reach a significant resolving power. Even though this quantification of the critical number of molecules only offers an approximate value, it is still a far more rigorous mathematical description of the effect of labeling density on resolution than that offered by the Nyquist sampling theorem. It is worth noting that for the *d/sd* = 2, we made

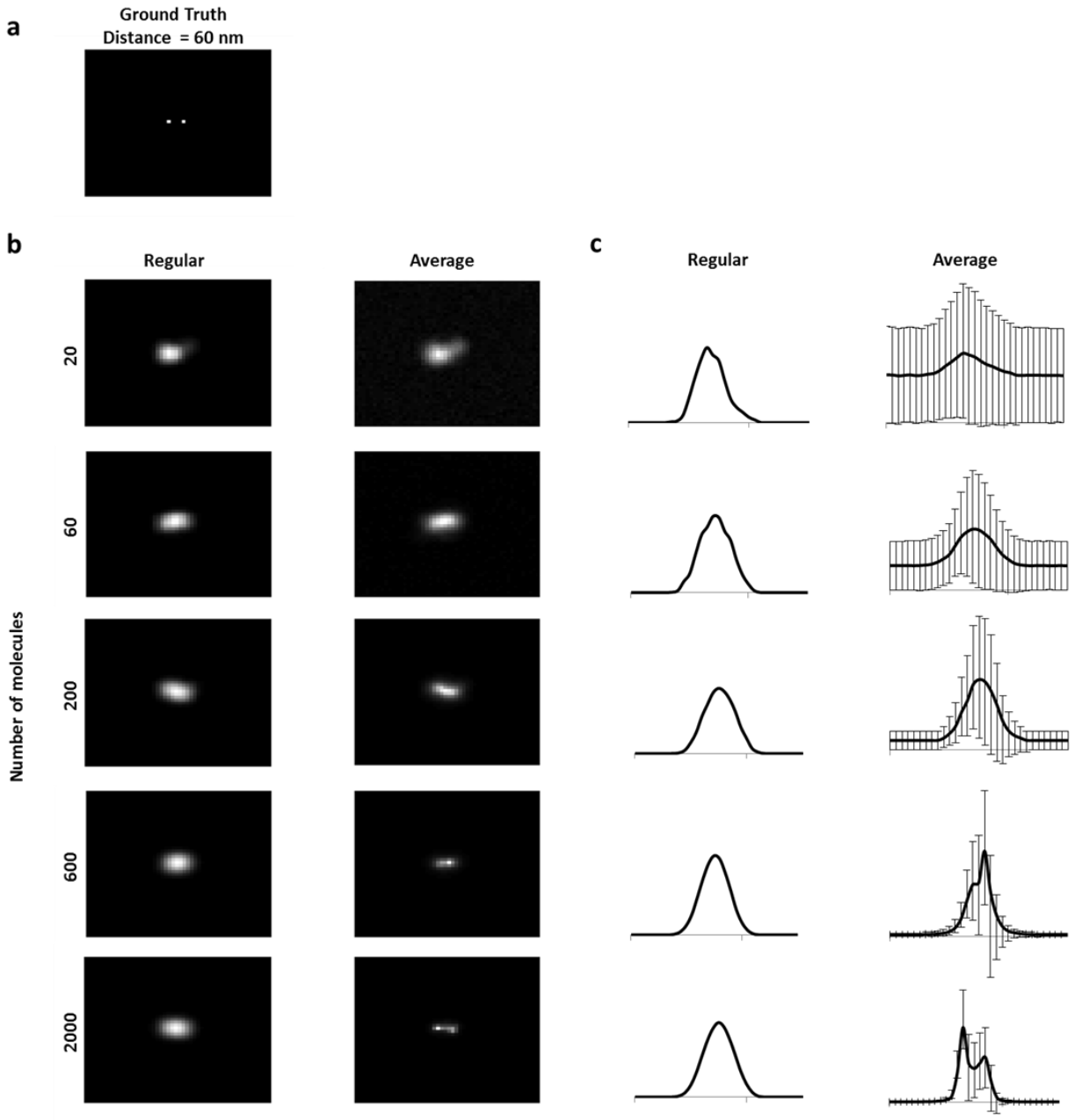

Fig. 3.9. Analysis of two point sources 60 nmapart in a 2D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Images are 0.75x0.75 um^2 and error bars represent standard deviation.

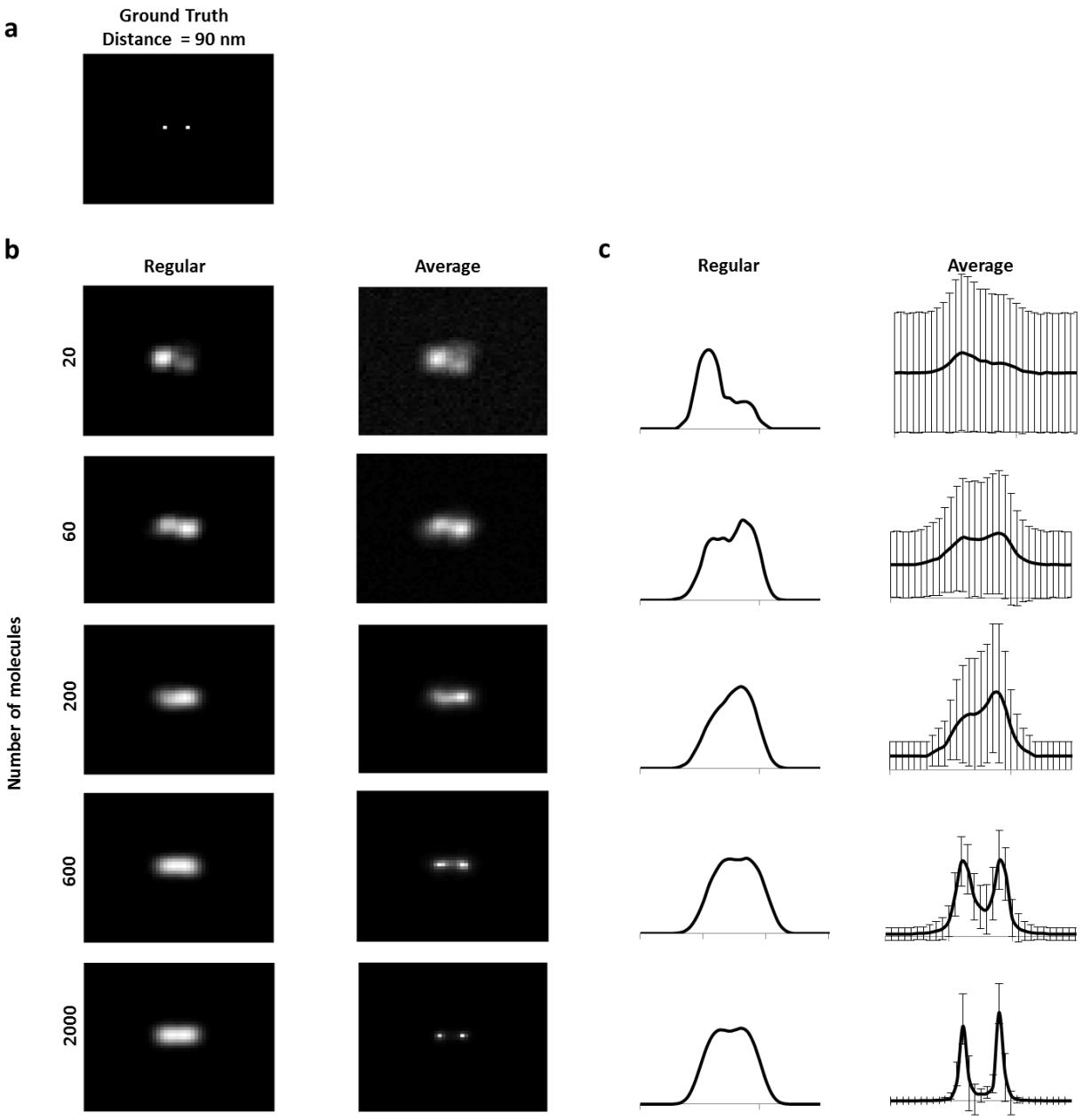

Fig. 3.10. Analysis of two point sources 90 nm apart in a 2D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Images are 0.75x0.75 um^2 and error bars represent standard deviation.

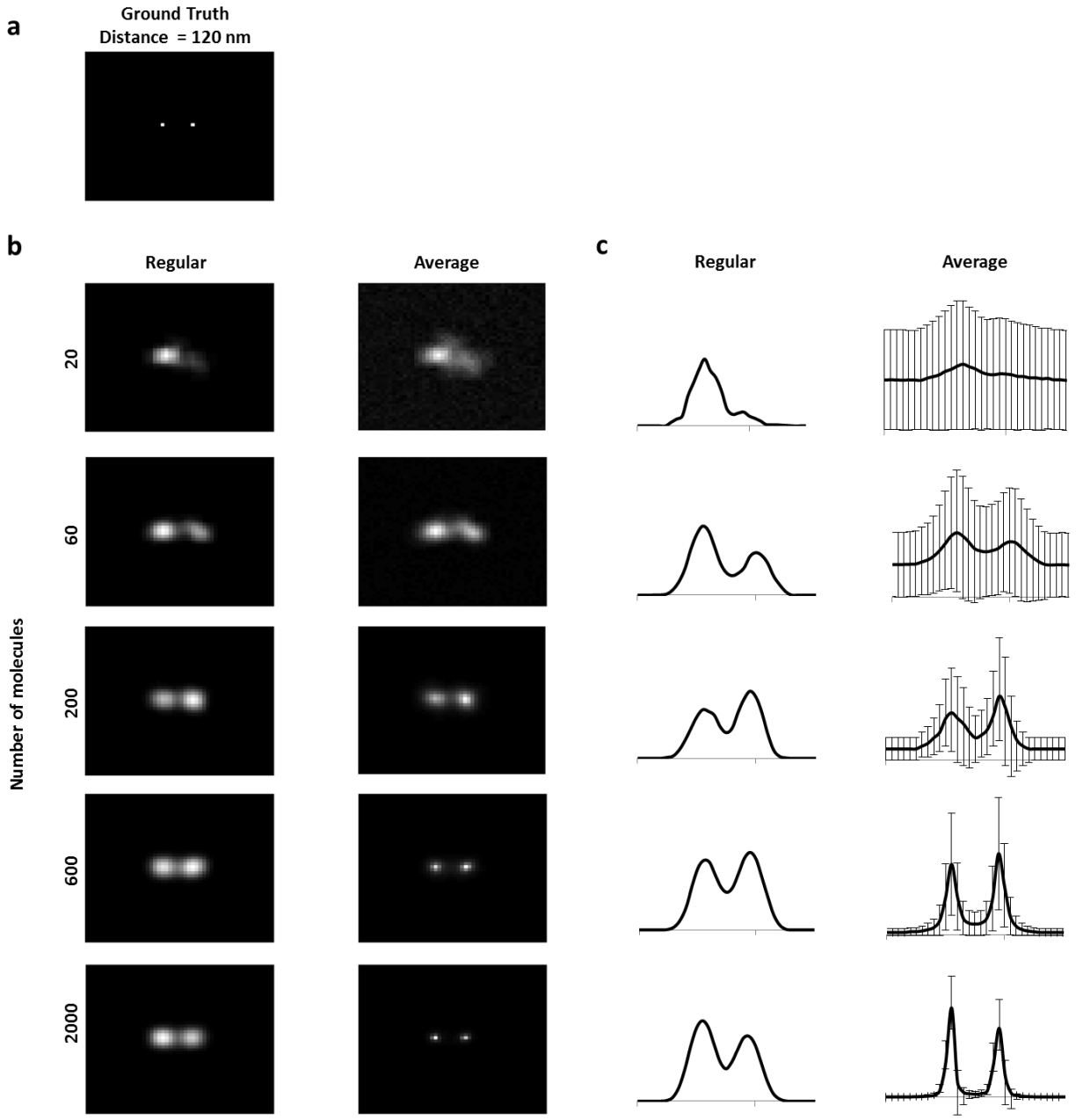

Fig. 3.11. Analysis of two point sources 120 nm apart in a 2D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Images are 0.75x0.75 um^2 and error bars represent standard deviation.

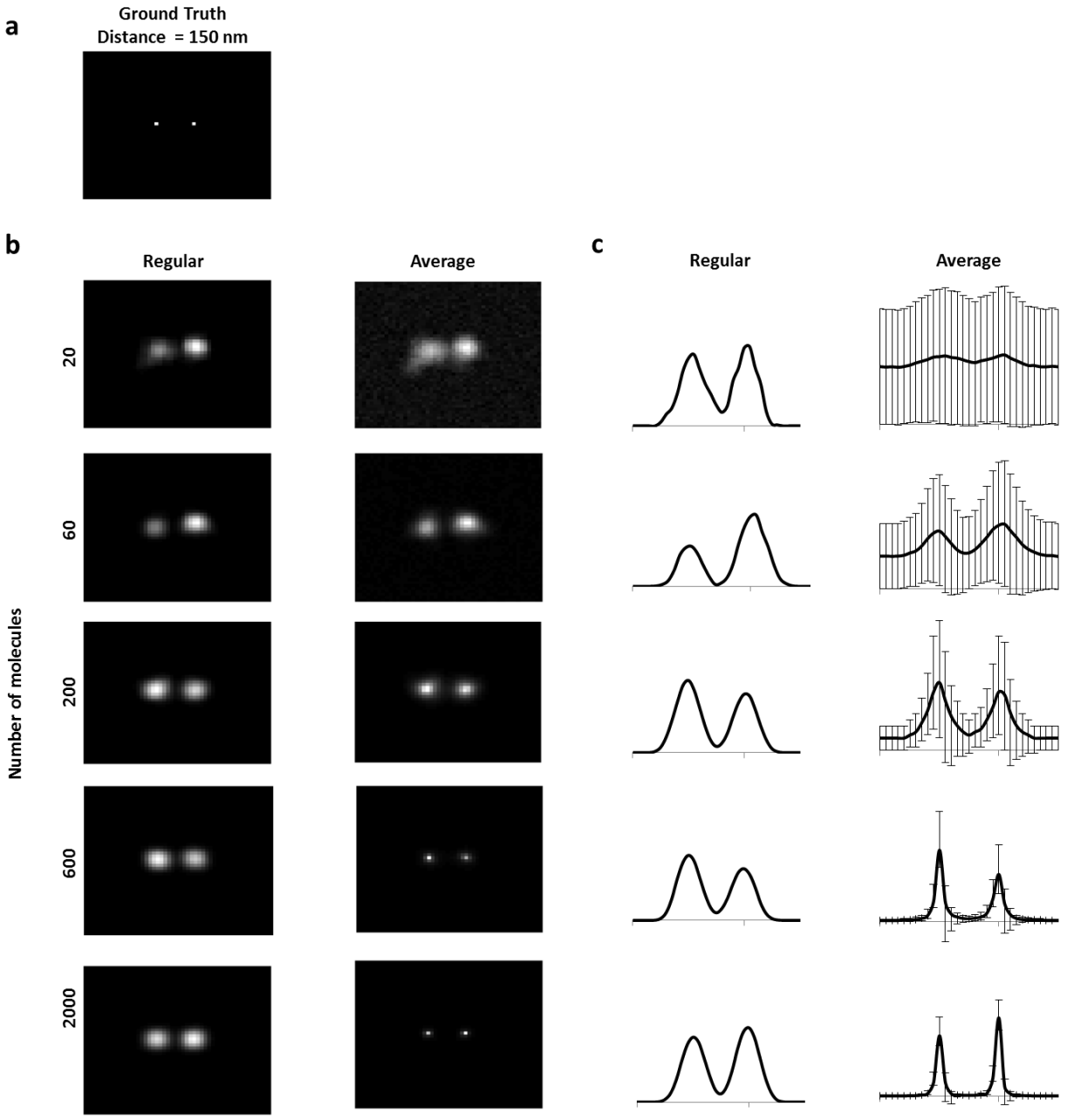

Fig. 3.12. Analysis of two point sources 150 nm apart in a 2D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Images are 0.75x0.75 um^2 and error bars represent standard deviation.

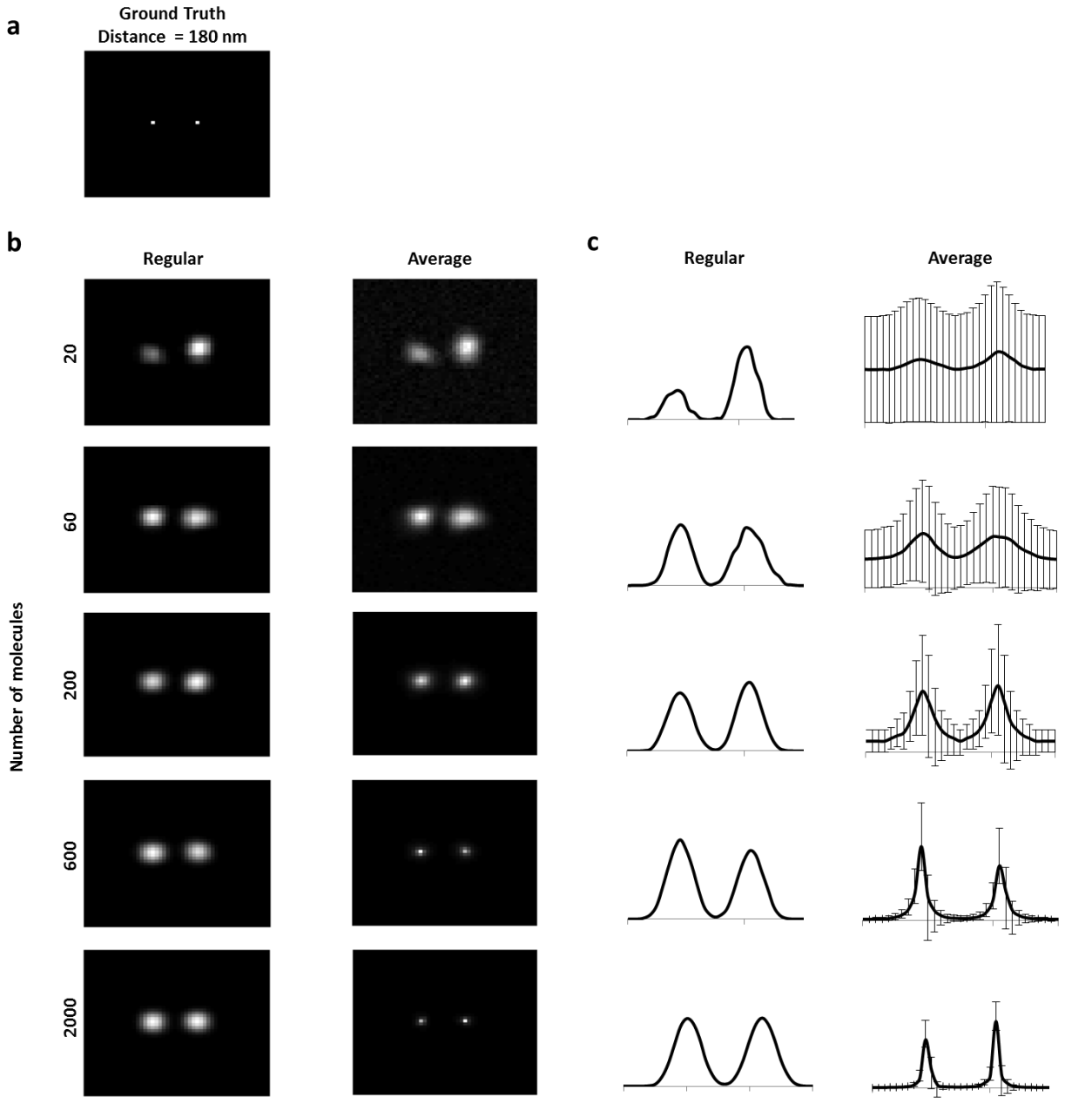

Fig. 3.13. Analysis of two point sources 180 nm apart in a 2D image. (a) The ground truth showing the two point sources used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (20, 60, 200, 600, 2000). (c) Linescans corresponding to the images in (b). Images are 0.75x0.75 um^2 and error bars represent standard deviation.

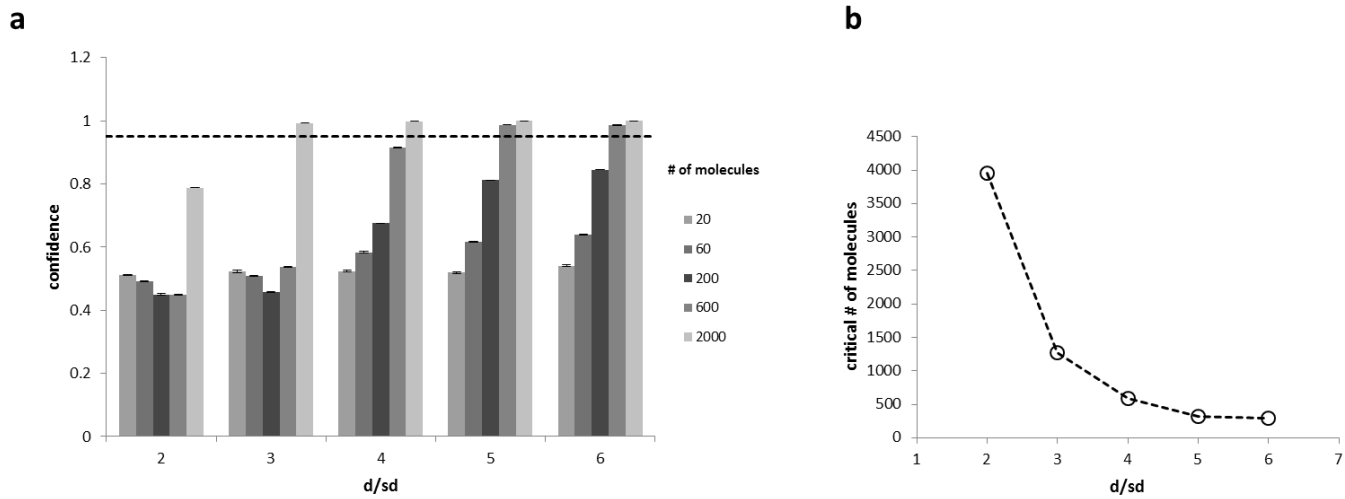

Fig. 3.14. Summary of the 2D simulation results for the two point sources. (a) The bar plots show the confidence in resolving the two features with the increasing number of molecules for each of the *d/sd* (distance to standard deviation ratio) scenarios. Dashed line represents the 95% confidence level and error bars represent the s.e.m, n = 3. (b) shows how the critical number of molecules required for resolving the two features converges with the increase in *d/sd* values.

an estimate for the critical number of molecules by extrapolation rather than interpolation since none of the simulated number of molecules reached the 95% confidence level.

The next step was to explore structures other than the simple two point sources so we explored the structure of two parallel filaments, which might represent a cytoskeletal structure (Fig. 3.15-3.17). Simulations were done using ground truth images of two parallel filaments (each is 300 nm in length) with varying distances between them (90, 120 and 150 nm). We took single vertical linescans that cut through the middle of the filaments to assess the confidence in resolution. Again, we see the same behavior of the improved resolution of the averaged images as the number of molecules increases from 400 to 40000 molecules or the *d/sd* ratio increases from 3 to 5. The effect of the number of molecules on the resolution confidence is summarized in Fig. 3.18a. As expected, we saw the same convergence of the critical number of molecules required to reach the 95% confidence level even at very high separation between features (Fig. 3.18b).

The structure of the parallel filaments allows us to explore another factor that might affect our confidence in resolution, which is the prior knowledge of the underlying structure. Because we know that the two lines are perfectly parallel, i.e.: all pixels within each filament are equivalent and all pixels in the trough valley between the two filaments are also equivalent, we can bin all these pixels together to get a pseudo 1D structure. This binning dilutes away the uncertainties from single pixels and increases the confidence in resolution. To illustrate this we compared the binned linescans with the single linescans and indeed, the binned scans gave consistently higher confidence values than the single scans that even fail to reach significance in case of $d/sd = 3$ (Fig. 3.18b). We can use the same binning principle to average between structures that our prior knowledge tells us are similar. Focal adhesions are one example of such structures where we can

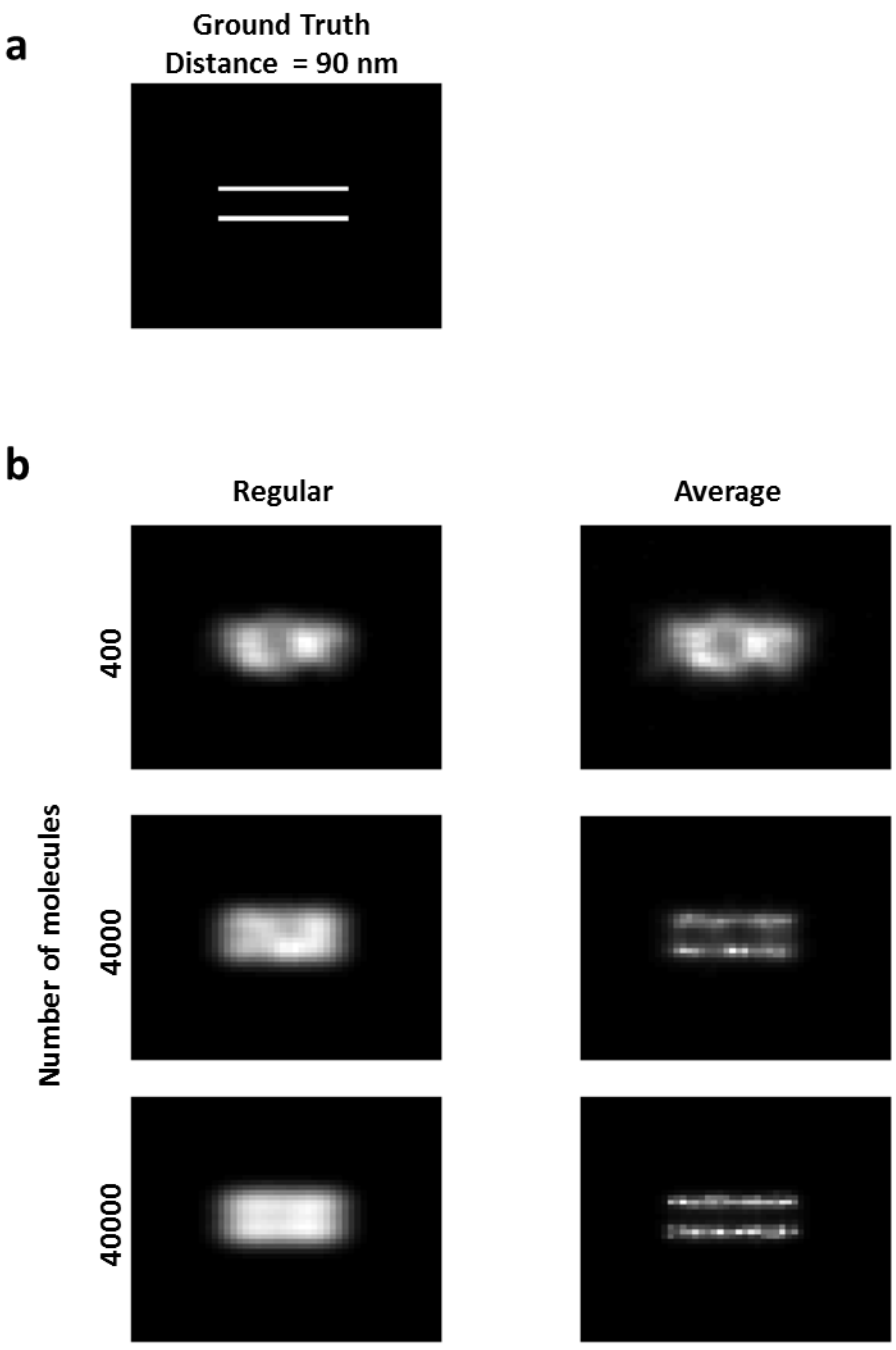

Fig. 3.15. Analysis of two parallel filaments 90 nm apart in a 2D image. (a) The ground truth showing the two parallel filaments used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (400, 4000 and 40000). Images are 0.75x0.75 um^2.

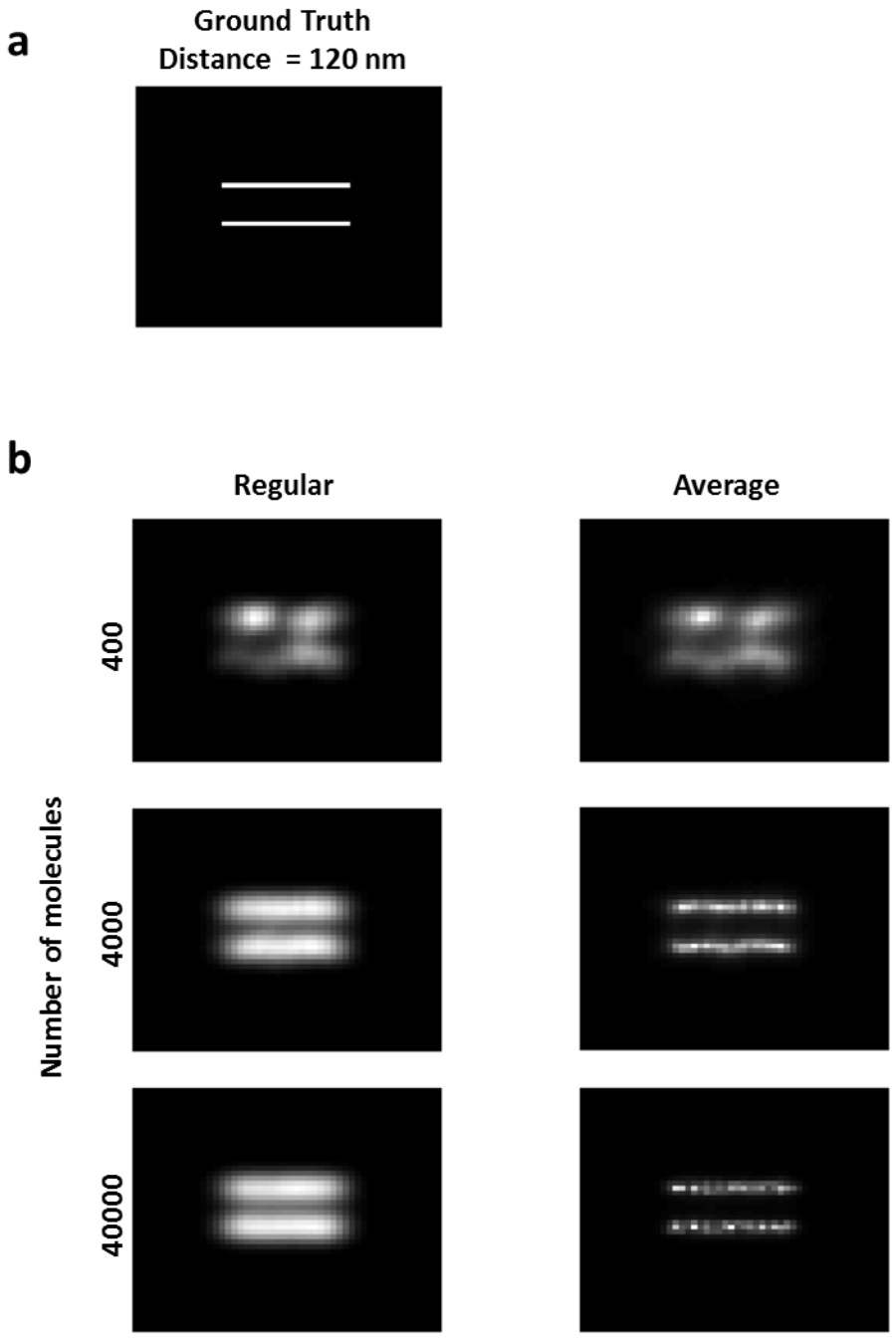

Fig. 3.16. Analysis of two parallel filaments 120 nm apart in a 2D image. (a) The ground truth showing the two parallel filaments used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (400, 4000 and 40000). Images are 0.75x0.75 um^2.

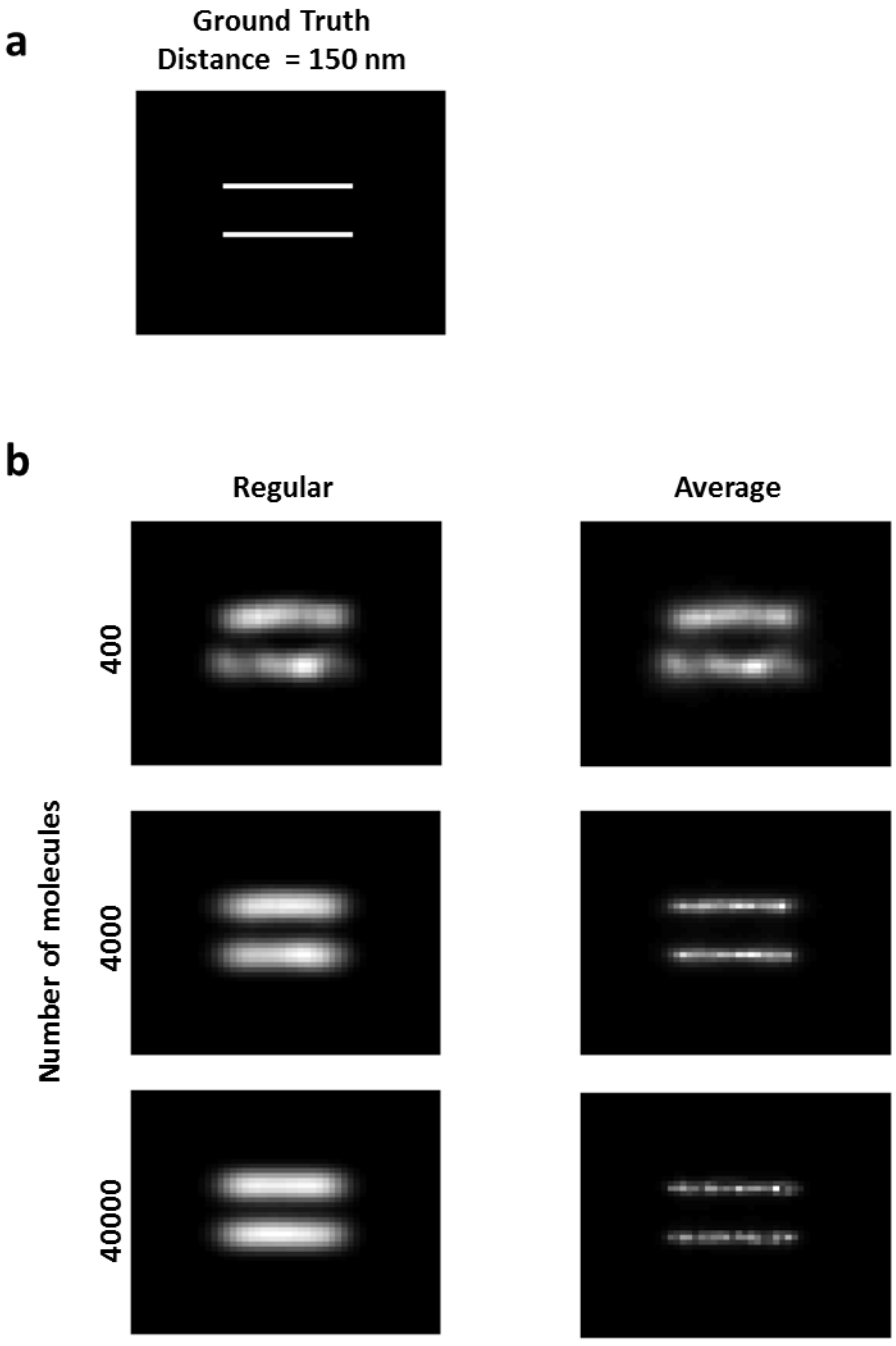

Fig. 3.17. Analysis of two parallel filaments 150 nm apart in a 2D image. (a) The ground truth showing the two parallel filaments used for generating simulated localizations. (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (400, 4000 and 40000). Images are 0.75x0.75 um^2.

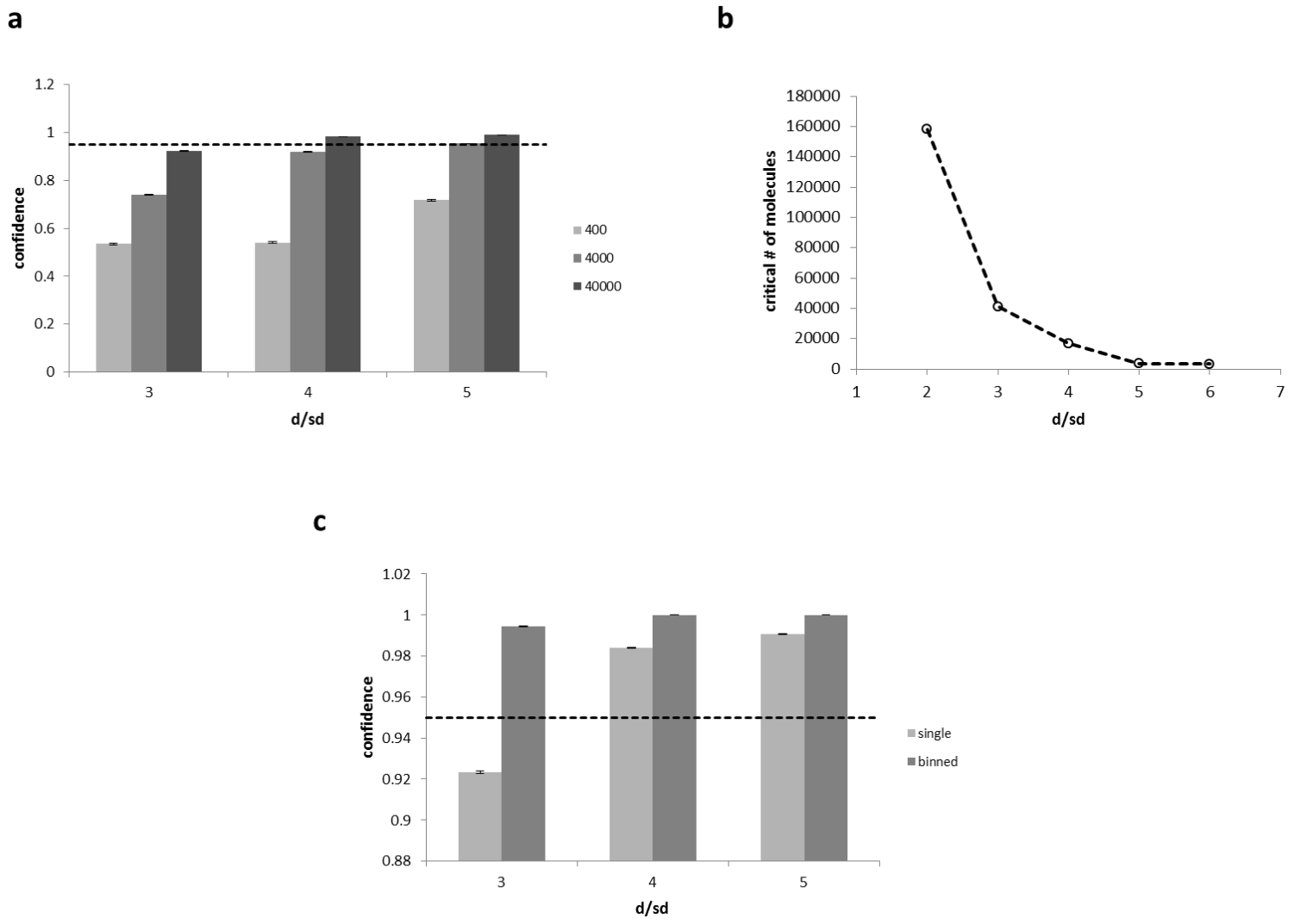

Fig. 3.18. Summary of the 2D simulation results for the two parallel filaments. (a) The bar plots show the confidence in resolving the two features with the increasing number of molecules for each of the *d/sd* (distance to standard deviation ratio) scenarios. (b) shows how the critical number of molecules required for resolving the two parallel features converges with the increase in *d/sd* values. (c) shows the comparison between the confidence values of binned and single linescans. Dashed lines represent the 95% confidence level and error bars represent the s.e.m, n = 3.

average the information from different adhesions to increase our confidence in the underlying spatial distribution of adhesion molecules.

To further emphasize our algorithm's ability to incorporate the effect of labeling density on resolution, we revisited the simulation results in the first row of Fig. 3.16b. We notice that the upper filament appears to be discontinuous, which we know is false. This kind of misleading information results from low number of detected molecules (low labeling density). If we take linescans through that filament in the regularly rendered super-resolution image and the average image generated by our algorithm, both images would look similar showing two peaks and a dip in-between but what our algorithm would also show is the very large uncertainty in these scans resulting from poor labeling density and represented by standard deviation error bars (Fig. 3.19). This is telling us that there isn't enough information to confidently say that these two peaks are resolved and this feature is of course lacking in regular rendering algorithms.

Finally, we wanted to demonstrate the ability of our algorithm to resolve even more complicated structures to assess its ability to perform on biological samples. We did that by generating a ground truth image of the letters "CCAM", which is an abbreviation of the "Center for Cell Analysis and Modeling", and then we ran simulations with an increasing number of molecules (1000, 10000 and 100000 molecules) (Fig. 3.20). Consistent with our earlier assessment, our algorithm does perform very well on this complicated structure as evident from the improved resolution of the averaged image in comparison to the regularly rendered image as the number of molecules increases. This was quantified by computing the correlation coefficients between these images and the ground truth image (Fig. 3.20c).

As mentioned in the methods section, we evaluated how using a varying sd would affect the

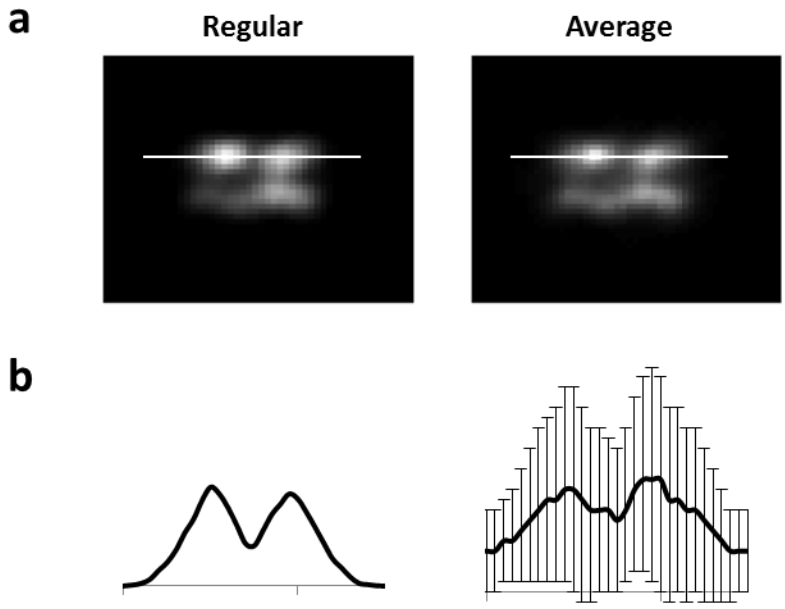

Fig. 3.19. Analyzing false discontinuous structures. (a) Comparison between the regular image rendering (left) and the averaged image generated by Gibbs sampling (right). The images were generated using 400 molecules and a separation of 120 nm between filaments. (b) shows the linescan results from the corresponding images in (a). Error bars represent standard deviation.

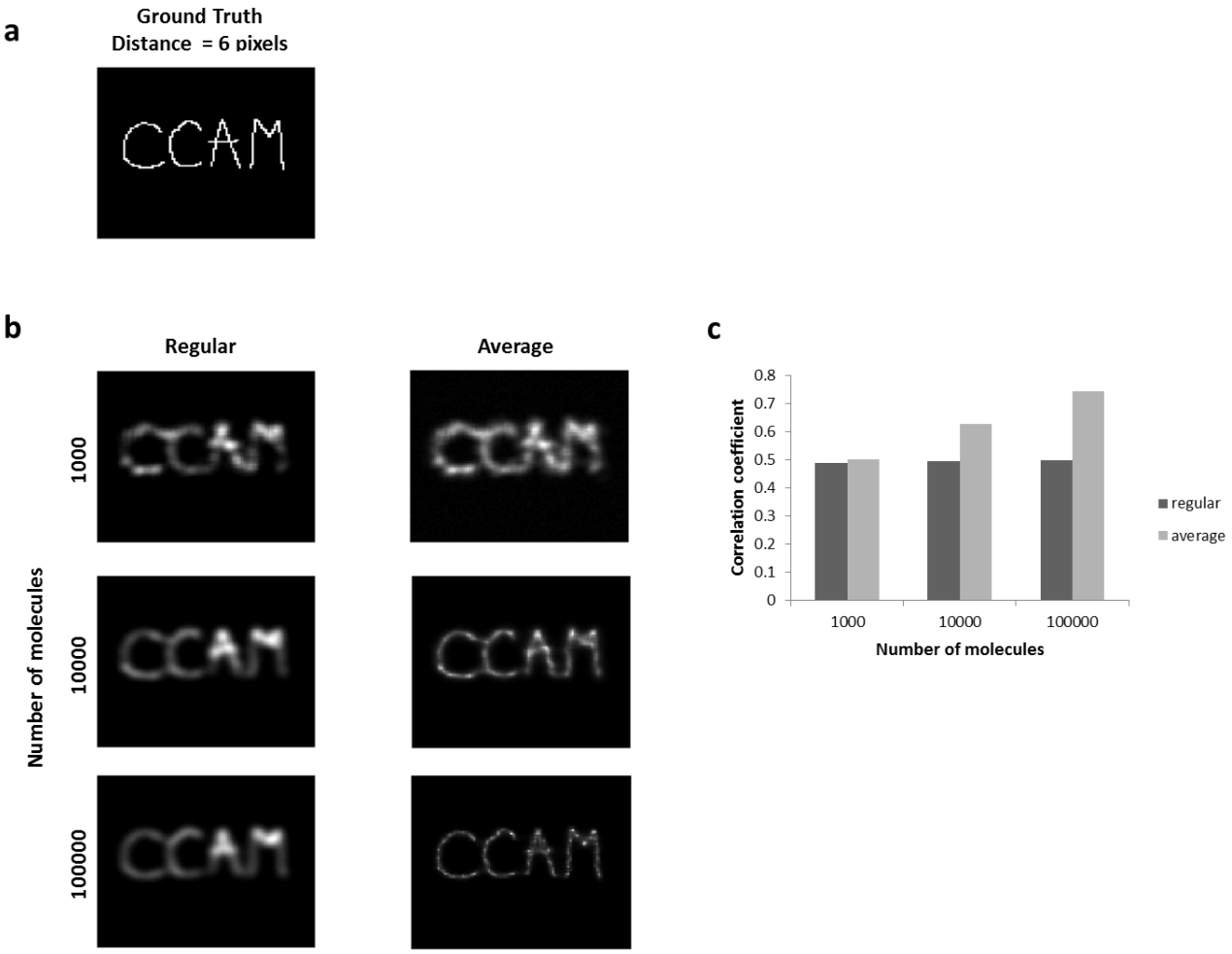

Fig. 3.20. Assessing the algorithm performance against a complex structure. (a) The ground truth showing the word "CCAM". (b) Comparison between the regular image rendering algorithm (left) and the average of the samples generated by Gibbs sampling (right) for differet number of molecules (1000, 10000 and 100000). Images are 0.75x0.75 um^2. (c) Corrlation coefficients of each of the images in (b) to show how correlated they are to the ground truth image in (a).

simulation results in comparison to the fixed sd assumption that we used throughout this chapter. To do that we compared the results from the 2D simulations of 2 point sources separated by 120 nm simulated under the fixed 30 nm sd assumption to the corresponding scenario of varying sd where each localization is given a different sd value that ranged between 15 and 45 nm. The results showed no significant difference between both scenarios (Fig.3.21).

3.4.2 Analyzing uncertainty in cluster size measurement

Nanocluster formation has been shown to mediate a variety of cellular functions as cell migration, proliferation, immune response and pathogen recognition (Garcia-parajo et al. 2014). Measuring these cluster sizes has been used to assert mechanistic assumptions or to explain cellular behavior (Mizuno et al. 2016; Bar-On et al. 2012; Smith and Verkman 2015). That's why a lot of effort has been undertaken to develop methods to estimate cluster sizes and to differentiate between real and apparent clusters that might result from localization uncertainty as we previously explained (Baumgart et al. 2016; Veatch et al. 2012). Unfortunately, the available methods fail to quantify the uncertainty in cluster size measurement as a function of localization uncertainty and labeling density, which might lead to inferring the wrong assumptions about the system under study.

To test our algorithm performance in giving estimates on cluster size uncertainty, we simulated a series of ground truth images with randomly distributed clusters of varying diameters from 60 to 150 nm (Fig. 3.22a-3.25a). We then generated single molecule localization data from 25000, 50000 and 100000 molecules for each scenario keeping the localization uncertainty standard deviation fixed at 30 nm. Finally, we compared the regularly rendered images to those

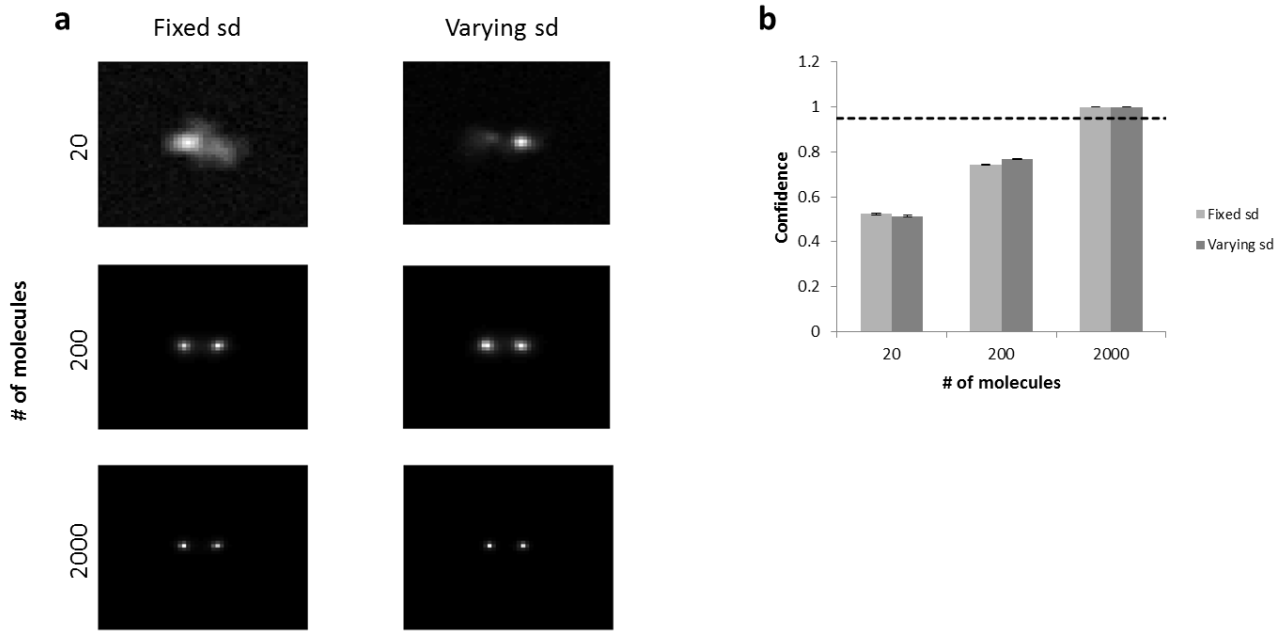

Fig. 3.21. Evaluating difference between the fixed and varying sd assumptions on simulation results. (a) The averaged images of samples obtained by the Gibbs sampler using under the fixed sd (left) and the varying sd (right) assumptions. Each row represents a different number of simulated molecules being localized. (b) Comparison between the confidence levels obtained using either assumption from (a) for the different number of molecules. Images are 0.75x0.75 um^2 and error bars represent standard deviation.

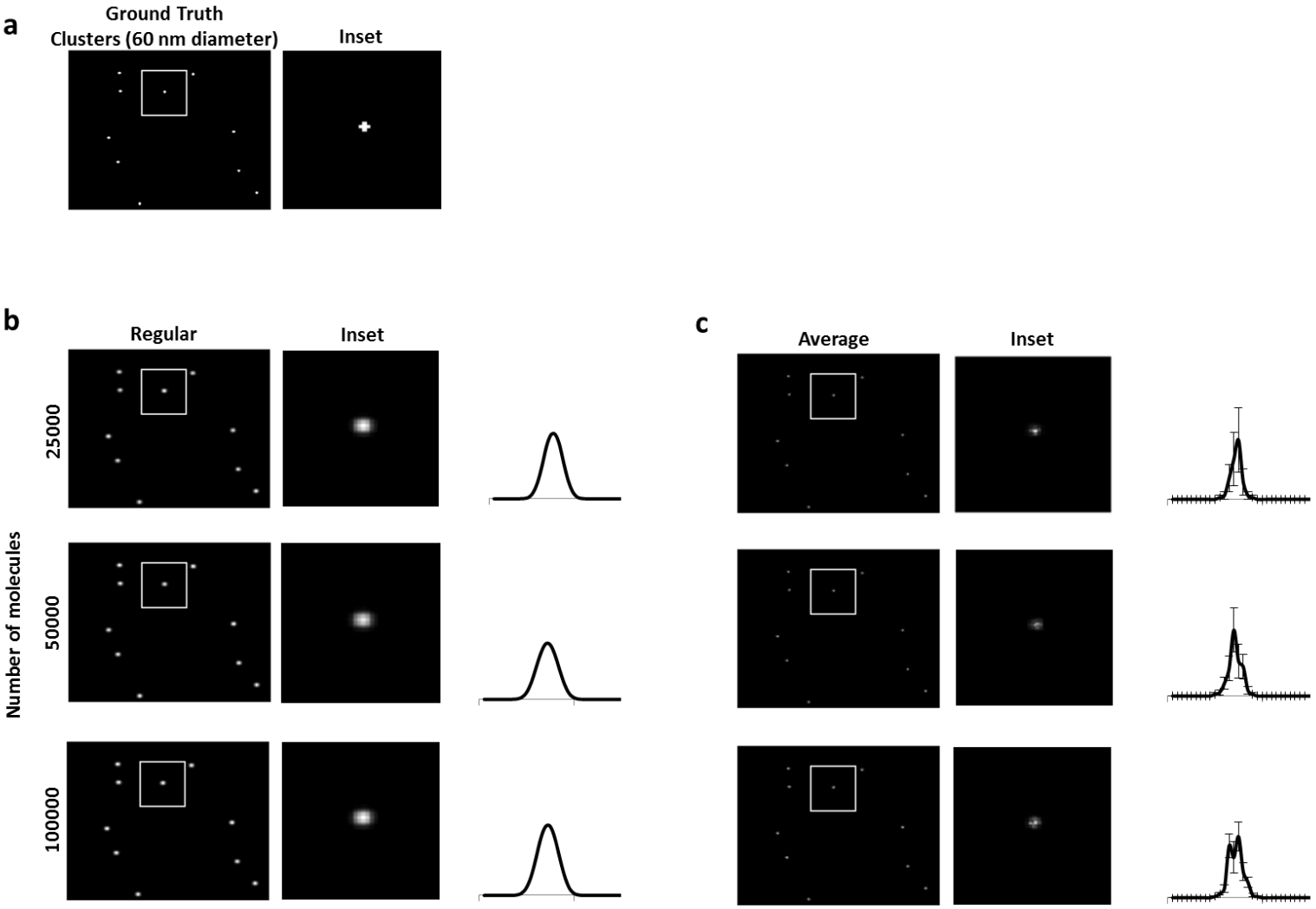

Fig. 3.22. Comparison between regular image rendering and statistical average images for clusters 60 nm in diameter. (a) Ground truth image of 10 randomly distributed clusters with 60 nm diameters. (b) shows the full regularly rendered images (left), zoomed in insets on one of the clusters (middle) and linescans through the clusters (right) for 25000, 50000 and 100000 molecules. (c) shows the full statistical averaged images (left), zoomed in insets on one of clusters (middle) and linescans through these clusters (right) for the same number of molecules as the regular rendering. Error bars represent standard deviation.

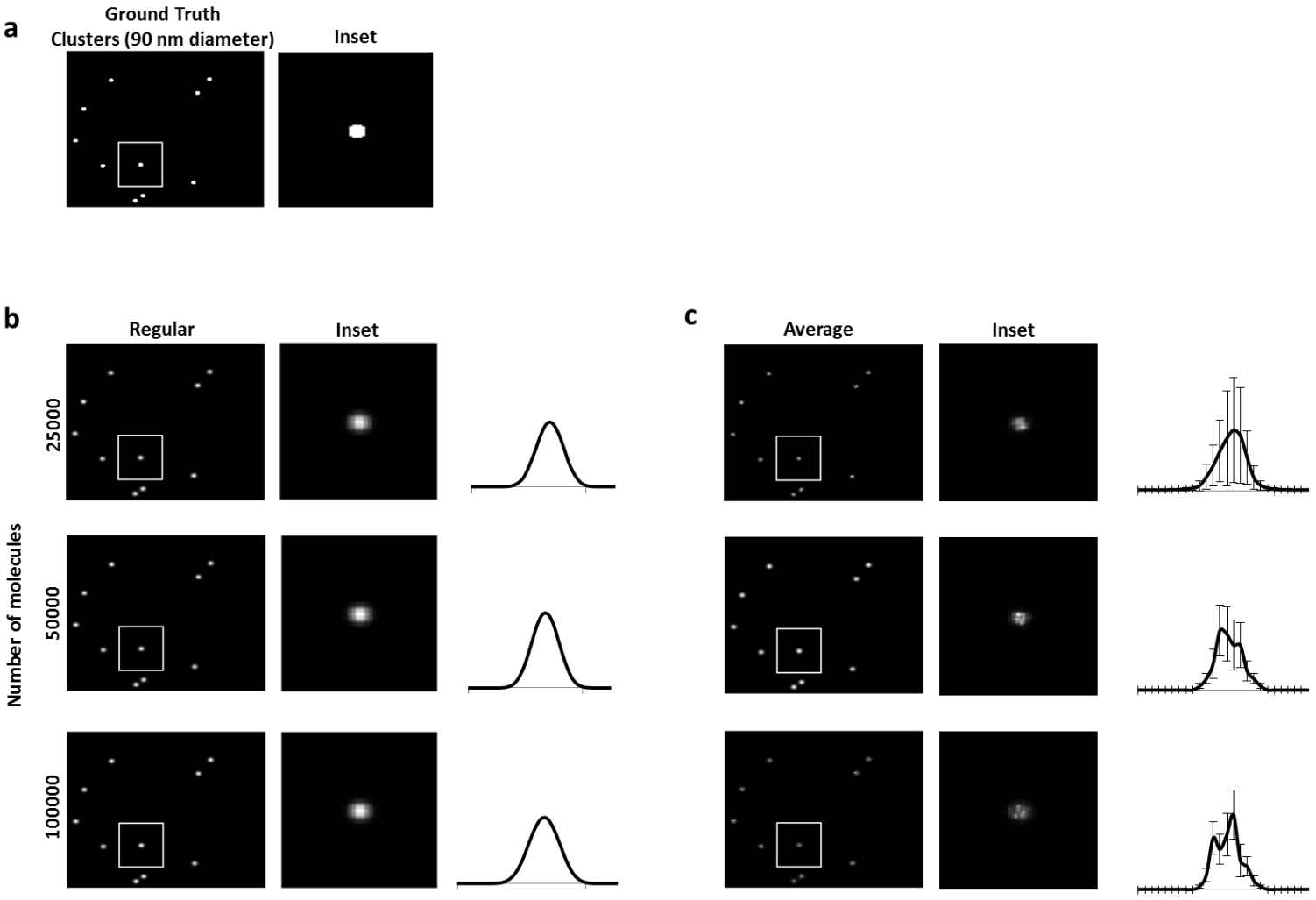

Fig. 3.23. Comparison between regular image rendering and statistical average images for clusters 90 nm in diameter. (a) Ground truth image of 10 randomly distributed clusters with 60 nm diameters. (b) shows the full regularly rendered images (left), zoomed in insets on one of the clusters (middle) and linescans through the clusters (right) for 25000, 50000 and 100000 molecules. (c) shows the full statistical averaged images (left), zoomed in insets on one of clusters (middle) and linescans through these clusters (right) for the same number of molecules as the regular rendering. Error bars represent standard deviation.

\

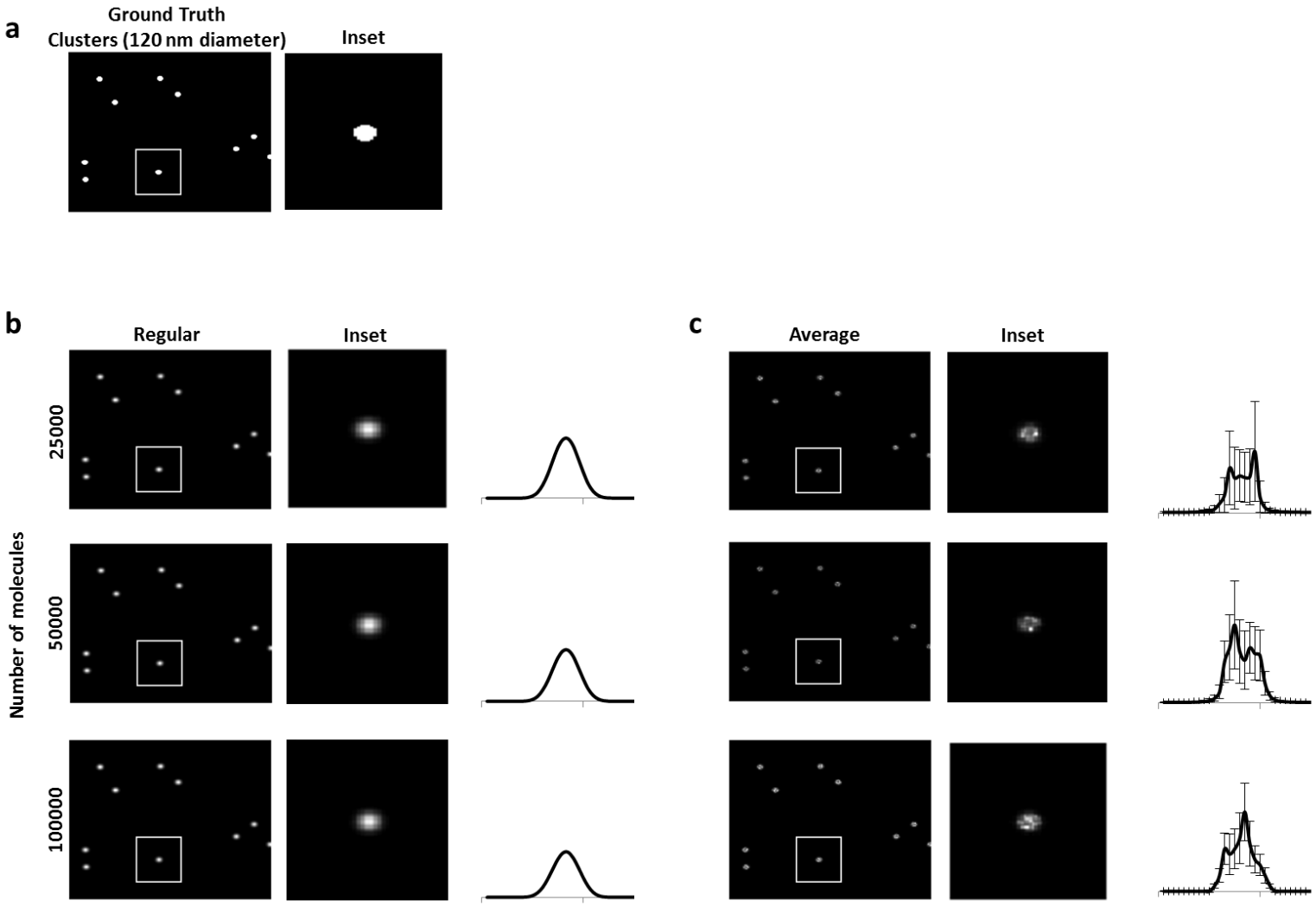

Fig. 3.24. Comparison between regular image rendering and statistical average images for clusters 120 nm in diameter. (a) Ground truth image of 10 randomly distributed clusters with 60 nm diameters. (b) shows the full regularly rendered images (left), zoomed in insets on one of the clusters (middle) and linescans through the clusters (right) for 25000, 50000 and 100000 molecules. (c) shows the full statistical averaged images (left), zoomed in insets on one of clusters (middle) and linescans through these clusters (right) for the same number of molecules as the regular rendering. Error bars represent standard deviation.

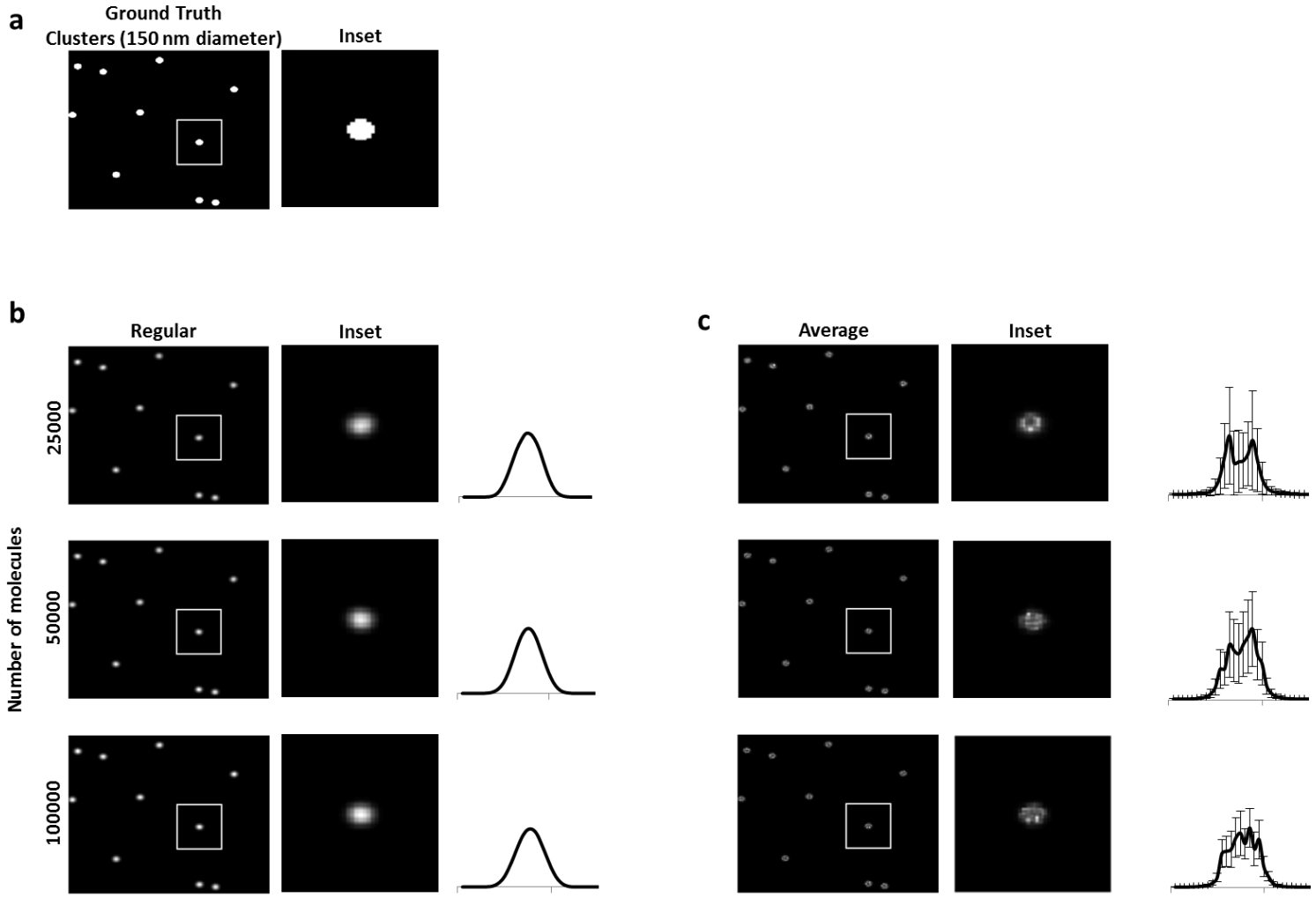

Fig. 3.25. Comparison between regular image rendering and statistical average images for clusters 150 nm in diameter. (a) Ground truth image of 10 randomly distributed clusters with 60 nm diameters. (b) shows the full regularly rendered images (left), zoomed in insets on one of the clusters (middle) and linescans through the clusters (right) for 25000, 50000 and 100000 molecules. (c) shows the full statistical averaged images (left), zoomed in insets on one of clusters (middle) and linescans through these clusters (right) for the same number of molecules as the regular rendering. Error bars represent standard deviation.

of the statistical average of the images generated by our Gibbs sampler after getting rid of the first 20000 burn-in samples. A general observation is that the regular image rendering algorithm gives clusters with larger sizes than the corresponding clusters from the averaged images, which tend to be closer to the ground truth size (Fig. 3.22b&c-3.25b&c). This is expected from the regular rendering algorithm since oversampling keeps adding localizations that are offset from the true localizations inside the cluster and hence, the resulting cluster will appear bigger than it really is. Taking linescans through the clusters in both cases again reveals the advantage of our algorithm that shows the uncertainty in the intensity values for every pixel in the scanned line and we can see that the more molecules we detect the more certain we are about these values (Fig3.22b&c-3.25b&c).

We also simulated randomly scattered molecules using a ground truth image with random point sources that we used to generate the same number of localizations as with the cluster scenarios (Fig. 3.26). The apparent cluster problem is clear in the regularly rendered images where single point sources generate cluster-like morphologies as a result of measurement noise with even some cases where two point sources collapse into one (Fig. 3.26a). On the other hand, the averaged images show enhanced resolution that converges to the ground truth with increasing the number of molecules and hence, the apparent cluster morphology gets diluted (Fig. 3.26b).

The samples generated by our Gibbs sampler provide us with the opportunity to explore the uncertainty in the cluster size measurement by analyzing the cluster size distributions among the samples. We used thresholding to get rid of background noise and to compute cluster diameters (see methods). The cluster size distributions are presented in Fig. 3.27a-d by boxplots that show the median, first and third quartiles and the ranges (excluding outliers) of 1000 samples generated by our algorithm for each of the clustered scenarios. The distributions show a little

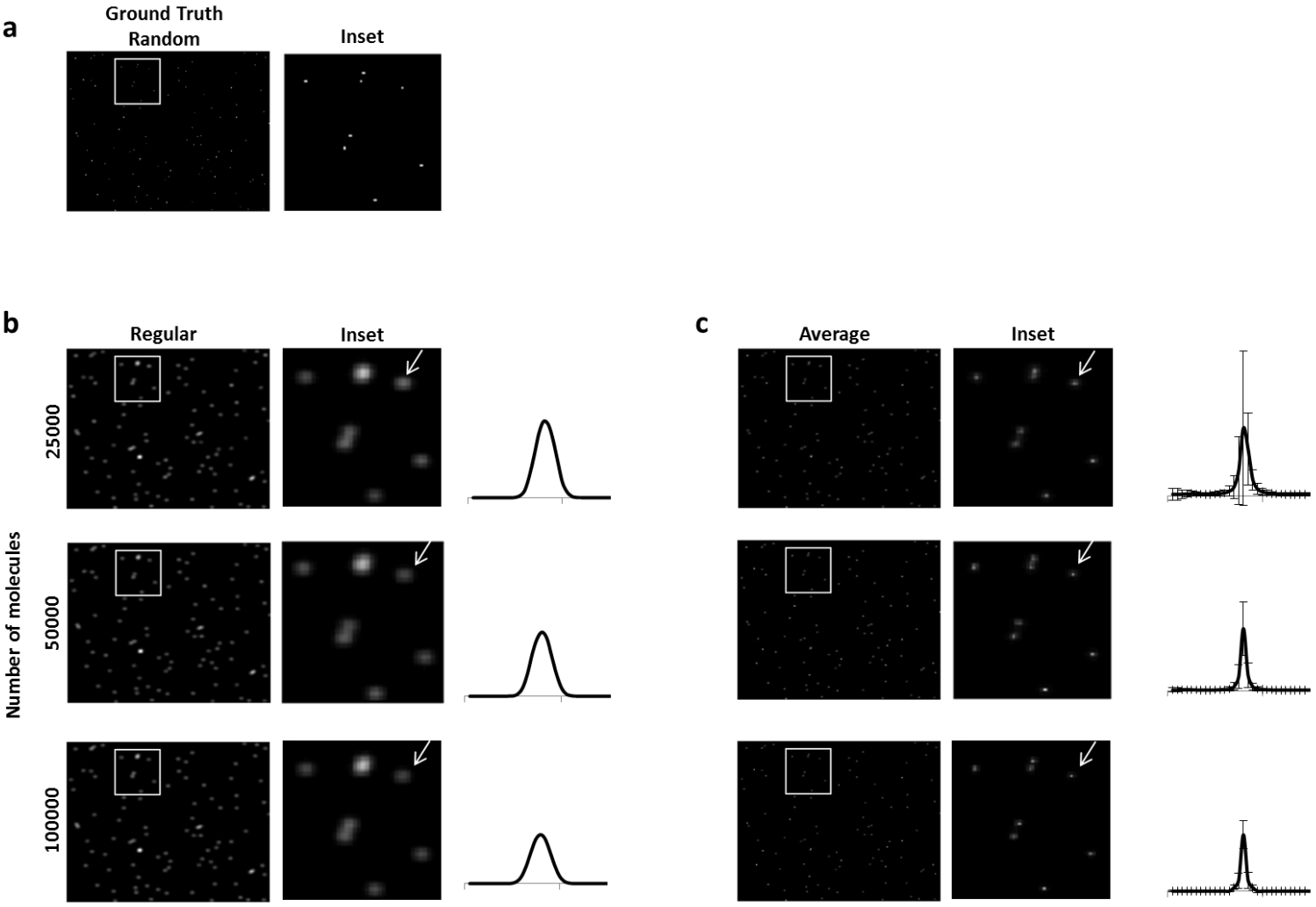

Fig. 3.26. Comparison between regular image rendering and statistical average images for randomly distributed molecules. (a) Ground truth image of 100 randomly distributed point sources. (b) shows the full regularly rendered images (left), zoomed in insets on part of the image (middle) and linescans through the structure pointed at by the white arrows (right) for 25000, 50000 and 100000 molecules. (c) shows the full statistical averaged images (left), zoomed in insets on part of the image (middle) and linescans through the structure pointed at by the white arrows (right) for the same number of molecules as the regular rendering. Error bars represent standard deviation.

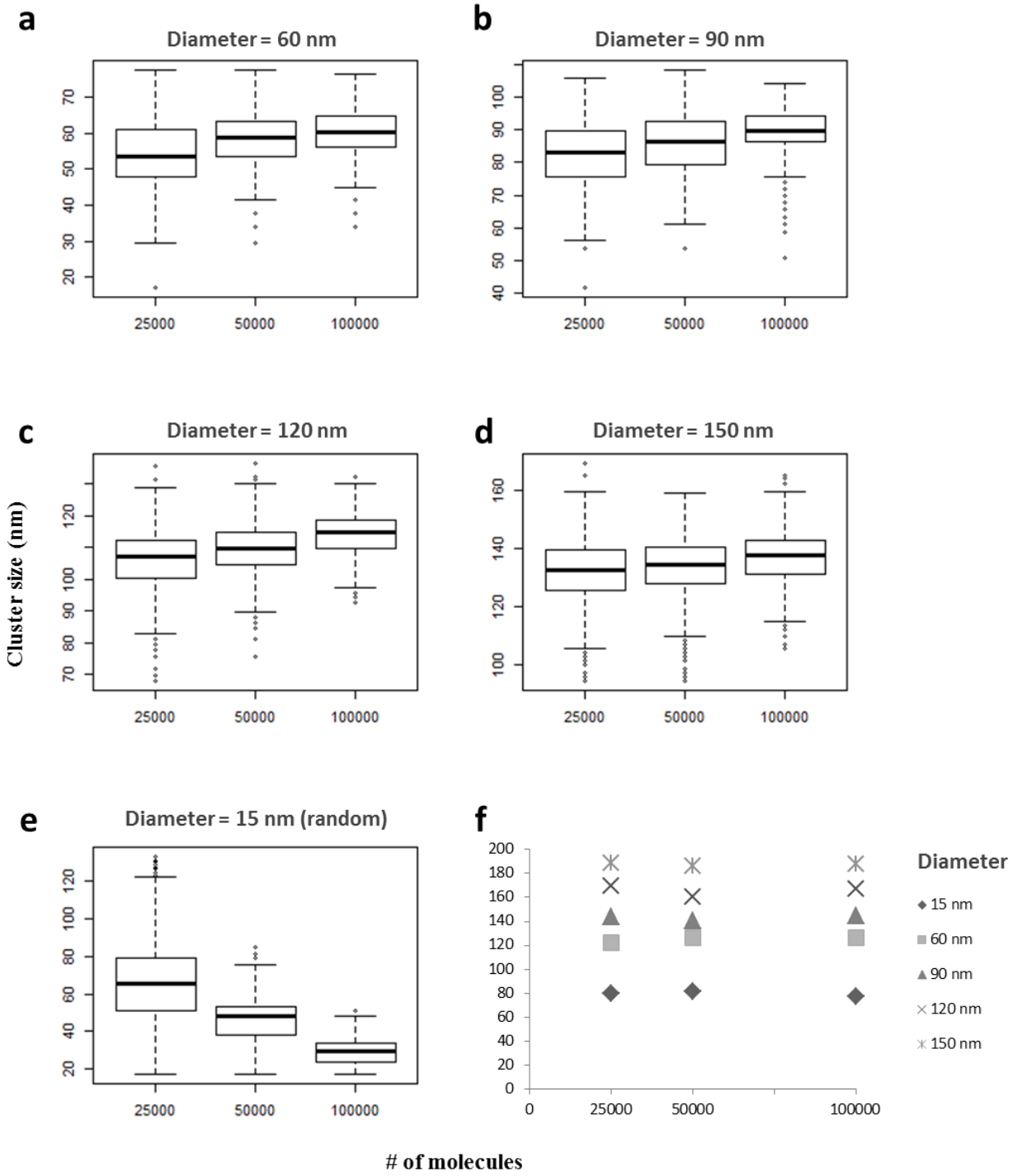

Fig. 3.27. Cluster size distribution analysis. The figure shows the cluster size distribution for ground truth cluster diameters of 60 (a), 90 (b), 120 (c) and 150 nm (d). (e) shows the size distribution for randomly distributed ground truth images. Boxplots represent the median, first and third quartiles (boxes) and ranges (whiskers) of all distributions and distributions are measured for the 25000, 50000 and 100000 molecules scenarios. (f) shows the cluster sizes measured from the regularly rendered images for the different ground truth cluster sizes and different numbers of molecules.

underestimation for the medians at lower molecular numbers that converge to the higher real cluster sizes as the number increases. An explanation for this initial underestimation is attributed to the thresholding method we are using to estimate cluster sizes since the image samples generated have pixels with a reduced overall intensity values, which is a reflection of the lower confidence. This in turn leads to filtering out some of the pixels that form the size of the cluster after thresholding. Another observation we see in the boxplots is that with increasing the cluster size, there is a systematic underestimation of the true cluster size even with a higher number of molecules. The thresholding method shortcoming we explained earlier might also explain this observation as with a larger cluster sizes, there is a higher probability of lower intensity pixels that can be filtered out with thresholding. To circumvent this problem, we can use other methods for cluster size estimation as autocorrelation but the take-home message from the boxplots is that the more molecules we detect, the closer we get to the true cluster size and the more certain we are about this measurement and this is evident through the decreasing range of the cluster size distributions. On the other hand, the regularly rendered images give cluster sizes that are typically larger than the true values and are not sensitive to the increasing number of molecules detected (Fig. 3.27f).

A similar boxplot was made for the randomly distributed molecules and we can see that at lower number of molecules, the distribution median is far from an individual pixel size (15 nm) whereas with an increasing number of molecules, the median converges to the true size of a pixel and the distribution range decreases (Fig. 3.27e). As for the regularly rendered image, the estimated size is far larger than a single pixel size of 15 nm, which is another indication of the apparent cluster morphology resulting from such algorithms (Fig. 3.27f).

3.5 Discussion

The current work introduces a novel algorithm based on Gibbs sampling, which for the first time is able to give a rigorous quantitative description of the influence of measurement noise on the resolution of super-resolution images rendered by SMLM as well as cluster size measurement. The algorithm was tested against computationally simulated data and was able to quantify the effect of parameters as localization uncertainty and labeling density on the resolving power between two features. Surprisingly, the influence of labeling density was more prominent than anticipated and it was shown that even with a very high separation between two point sources, a relatively large number of molecules is still required to reach a statistical confidence that they are separated. The logical explanation is that this large number of molecules is required to completely abrogate the possibility that the two features might actually be connected but the underlying connecting structure isn't fully decorated by the insufficient number of molecules. This conclusion was further confirmed when we simulated two parallel filaments with a low molecular number scenario where one of the filaments showed a false discontinuous morphology. Our algorithm was able to reveal the low confidence in that false morphology, which is something the available image rendering algorithms aren't able to do. Something to take away from these results is how challenging it might be for SMLM methods that use permanent labeling for decorating the features of interest as PALM, STORM or FPALM to reach confidence levels in resolving features or in asserting that a feature is discontinuous as these methods typically have lower labeling densities. On the other hand, PAINT relies on using diffuse probes and can therefore reach much higher labeling densities and hence, much higher confidence can be achieved.

Our algorithm also reveals how unsatisfactory current image rendering algorithms are in constructing the final super-resolution images and how inefficient it is to consider different localizations as separate events. Because our algorithm circumvents that by considering an underlying probability distribution where close localizations might be arising from the same molecule or position, it is able to alleviate the measurement noise in the rendered image. This is evident from the significantly improved resolution of the averaged images over images constructed by the regular image rendering algorithm.

As for cluster analysis, our algorithm proved efficient in quantifying the uncertainty in cluster size measurement. Again, the higher molecular number is shown to be crucial to converge to the true cluster size especially for larger clusters and to truly differentiate between real and apparent clusters. This conclusion suggests that extra scrutiny needs to be considered when measuring cluster sizes under a low labeling density experimental protocol.

In conclusion, we believe that our algorithm offers a significant advantage to cell biologists applying SMLM methods to study subcellular features at the nanometer level. It provides them with a statistical quantification of measurement uncertainties, which would eventually help them better analyze data and reach firmer conclusions. Furthermore, our algorithm proved to be superior over current regular image rendering algorithms in procuring significantly better resolved images.

Chapter 4

Discussion and Future Directions

The role of SMLM methods in the advancement of cell biology is undeniable. These methods allowed cell biologists to break the ~ 200 nm Abbe diffraction limit for optical fluorescence microscopy and decipher cellular structures at an unprecedented resolution of 10-30 nm (E. Betzig et al. 2006; Rust, Bates, and Zhuang 2006; Hess, Girirajan, and Mason 2006; Sharonov and Hochstrasser 2006). The advantage that these methods offer over other super-resolution fluorescence microscopy methods is their extremely simple implementation and they achieve single-molecule localization through diluting the fluorescence signal in the time domain by either allowing only a subset of the fluorophores to be fluorescent at a given time in the case of PALM/STORM/FPALM (E. Betzig et al. 2006; Rust, Bates, and Zhuang 2006; Hess, Girirajan, and Mason 2006), or by using a low concentration of diffuse fluorescent probes as in the case of PAINT (Sharonov and Hochstrasser 2006; Kiuchi et al. 2015). This of course requires acquiring thousands of images of the same sample to make sure that all underlying structures have been sufficiently labeled before constructing the final super-resolution image.

The significant potential advantages offered by SMLM methods begged for efficient solutions to the remaining challenges they face. Two of the most prominent of these challenges are: (i) the sample-drift and (ii) the measurement noise quantification problems. The sample drift problem arises from the natural phenomenon of mechanical drifting of the stage where the sample is placed on and because SMLM methods require the acquisition of thousands of single-molecule frames that might take a long time before constructing the final super-resolution image, the stage drifting during that acquisition could significantly decrease resolution (Fig. 1.2a) (Lee et al. 2012). On the other hand, the unique data acquisition process led to unconventional measurement noise characteristics that can be quite difficult to quantify. The measurement noise is connected to the number of photons being collected from the fluorophore during the

localization stage of single molecules and it also results in lower resolution of the final constructed image (Fig. 1.2b) (Waters 2009).

The current body of work is focused on addressing these two issues. In chapter 2, I have demonstrated a novel approach to solving the sample drift problem through the Bayesian inference framework. The motivation behind this approach was instigated when we were looking at the spatial organization of an interesting protein called CrkI. CrkI is an adaptor protein that has an SH2 (Src-homology 2) domain, which binds specifically to phosphotyrosine residues (Mayer, Hamaguchi, and Hanafusa 1988). During cell migration, CrkI gets recruited to membrane structures called focal adhesions, which are macromolecular complexes that help the cell to anchor itself to the extracellular matrix (Nagano et al. 2012; Zaidel-Bar et al. 2004).

We did PALM imaging on cells expressing mEos-CrkI but the final super-resolution images were blurred because of the sample drift that we weren't able to decide whether CrkI localizes smoothly over the whole adhesion structure or if it localizes to discrete subdomains (Fig. 2.8). Previously, the available options were either using fiducial markers or performing image correlation to estimate and correct for drift. The first option complicates the experimental procedure, while the second option proved to be unsatisfactory since it can only estimate drift with large time-delay and cannot estimate drift between individual frames. That's why we turned to a new innovative approach by using Bayesian statistics. Our method is a post-processing analysis that can be done using the acquired single-molecule data without further complicating the experimental procedure and the results showed a significant improvement over the currently available image correlation methods. Using our algorithm we were able to see a clear punctate-like distribution of Crk across the focal adhesion structure (Fig. 2.8).

Given its success in correcting for drift and improving the resolution of the final super-resolution image of focal adhesion structures, which are relatively large structures (about 5 um in length and 1 um in thickness), it is conceivable to conclude that our algorithm will be even more successful for more spatially refined structures as actin filaments, microtubules or nanoclusters. This was already established through simulated images but we anticipate the same success with real experimental data.

The established efficiency and generalizability of our drift correction algorithm proves that it is a significant addition to the fields of cell biology and SMLM. This significance was even more evident when soon after we published the algorithm, we were contacted by a researcher who downloaded it and had a question about implementation. He pointed out that his lab had been using fiducial markers to correct for drift but they wanted to switch to using our algorithm instead since it showed robustness and was much simpler to use.

A final remark on this part is that in all of our simulations we haven't accounted for photobleaching that might happen during acquisition in a real experiment. Nevertheless, we did simulate scenarios where low numbers of molecules are being detected per frame and our argument here is that if this number decreases further due to photobleaching at later frames, the data obtained from these frames would be useless anyway and will be discarded as they wouldn't add any valuable information about the drift or the underlying structure.

As for the future directions for the drift correction part of this thesis, we need to maintain our open-source algorithm that is made available to everyone on Github and to solve any bug that users might bring to our attention.

In chapter 3, I have demonstrated another algorithm based on Gibbs sampling that we developed to address the difficulties in quantifying measurement noise in localization microscopy. This problem might lead to localization errors because of the uncertainty in the localization precision (Thompson, Larson, and Webb 2002; Huang, Bates, and Zhuang 2010). Regular image rendering algorithms don't account for measurement noise and simply consider each detection as a separate event, which eventually results in a distorted final super-resolution image. Our algorithm overcomes this problem by considering close localization events correlated and imparts a probability that these events might actually be coming from the same molecule.

Since our algorithm is based on a sampling method (Gibbs sampling), a way to represent the images processed by the algorithm is to display the average image of all the samples collected after getting rid of the initial burn-in samples. The results show that these averaged images often have better resolution than regular rendering algorithms, especially when the labeling efficiency is high. This suggests that the conventional algorithms unnecessarily (and inaccurately) introduce blurring effects in the rendered image, which can be avoided using our sampling method. Another key innovation that our algorithm offers is that for the first time, we are able to use the samples generated to infer confidence statistics on the resolving power between features in a super-resolution image. To our knowledge, this feature is totally lacking in the current algorithms that analyze super-resolution images.

A surprising finding of our analysis was the fact that we need relatively large number of detected molecules to reach confidence levels that two features are resolved. This number ranged between 600 and 2000 molecules for the simulated two point sources scenarios and was much larger for more complex structures. This finding highlighted the need for more scrutiny when trying to infer statistical confidence or to build a complete structural map from super-resolution

images obtained by SMLM techniques that rely on permanent labeling as PALM, STORM or FPALM, where the typical labeling densities reached by these techniques are much lower than what might be required. On the other hand, PAINT depends on diffuse probes and can virtually reach any labeling density required, which makes it a much more suitable method to properly quantify statistical confidence in super-resolution images.

To demonstrate a real live application for our algorithm, I will go back to Crk1 localization PALM images I referred to earlier. After we corrected for drift and saw a punctate-like behavior, we came up with a hypothesis that Crk1 only gets recruited to regions under mechanical stress within the adhesion structure through a scaffold protein called p130Cas (Donato et al. 2010). This hypothesis was driven by previous work showing that p130Cas is a mechanosensor that primes its substrate domain for tyrosine phosphorylation and consequently, recruitment of Crk1 when it is under mechanical stress (Sawada et al. 2006). It was an attractive hypothesis but we were skeptical of the Crk1 localization images since we weren't sure whether the punctate-like distribution was real or if it is simply a result of undersampling that causes continuous structures to appear discontinuous. This need for a way to quantify resolution uncertainty was one of the major motivations for developing our algorithm that indeed gave pretty large uncertainties on the resolving power between these Crk1 subdomains and concluded that we don't have enough information to say for sure that the punctate-like distribution is the real morphology of Crk localization (Fig. 4.1).

In addition to the aforementioned application for our Gibbs sampling algorithm in analyzing the spatial distribution of focal adhesion proteins, other structures might be investigated such as cytoskeletal structures like actin filaments and microtubules and their associated proteins. Also

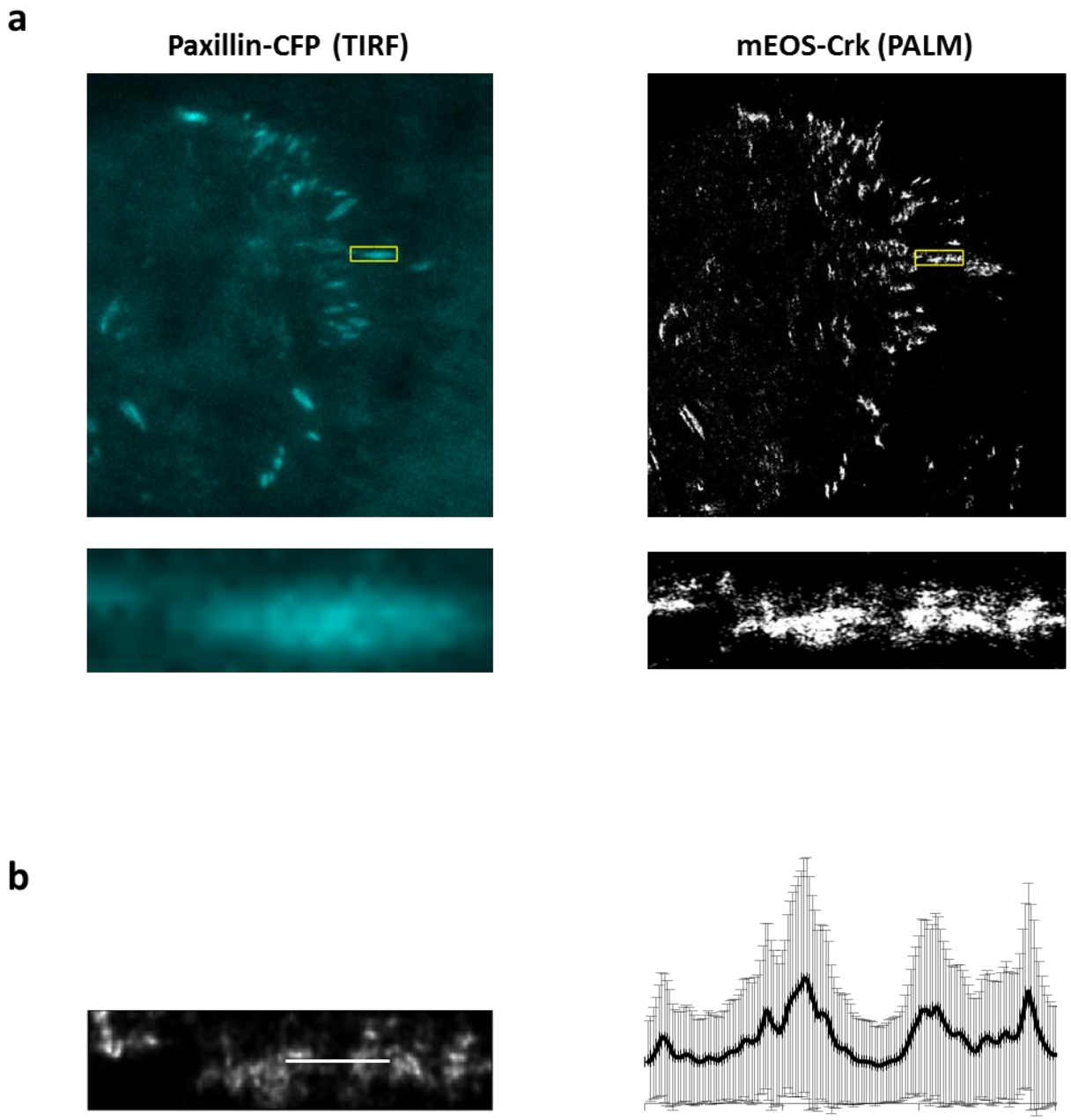

Fig. 4.1. Analysis of experimental Crk localization data with the Gibbs sampling algorithm. (a) TIRF image of paxillin used as a marker for focal adhesions (left) and the constructed PALM image of Crk (right). Below each image is the zoom in on a single adhesion marked by the yellow box. Yellow boxes and insets are 5 um in length. (b) The average image of the Gibbs samples collected (left) and the linescan intensity measure showing the uncertainty as standard deviation error bars (right).

the algorithm can be used to explore cluster size distributions of cluster-forming proteins where the size could have mechanistic and functional implications as in the case of the different cluster sizes of granuphilin, a Rab27 effector, that is dependent on whether they are associated with insulin granules or not (Mizuno et al. 2016). Furthermore, the flexibility of the algorithm to be extended to process 3D images adds attractive targets that can be explored by the algorithm like the centriole structure and its associated proteins. What is peculiar about centrioles is that they seem to show distinct subdomains that change in number according to the growth phase and that centriolar proteins localize either in dot-like structures or as rings (Sonnen et al. 2012). It has also been shown that during centriole duplication, a sub-diffraction gap is opened in the centriole ring structure (Mennella et al. 2012). All this makes the centriole system perfect for analysis by our algorithm where confidence estimates can be drawn on the centriolar proteins localizations or on the centriolar gap formation.

One of the concerns to point out about our Gibbs sampling algorithm include the fact that the computational time increases linearly with increasing the number of molecules detected. We partially addressed that concern by parallelizing the code and by binning localizations from the same coordinates together, which significantly improved the computation time. Nevertheless, there is still room for further improvement by translating the MATLAB® code into some other language that is not bound by a maximum number of nodes because of licensing issues like R or Python or to translate it to the much more efficient language, C++.

Another concern is that, once more, we haven't incorporated the effect of photobleaching in any of our simulations. The reason is that including photobleaching proved to be a much more mathematically and computationally challenging problem than anticipated but this is definitely one of the important future directions for this project.

The logical next step for this project is to test the Gibbs sampling algorithm performance against various experimental samples. It is worth mentioning that in order to be computationally efficient, the full super-resolution image obtained from real experiments will have to be fragmented into smaller pieces and each piece will be analyzed separately by the algorithm. We don't believe that this is a big issue since most of the questions that our algorithm could be used to answer don't need an analysis of the whole image. Even in the few cases where the analysis of the whole image is crucial, a computational cluster or a cloud facility can be used to carry out simultaneous runs of the pieces.

During the writing of this thesis, Stefan Hell's group published a paper introducing a new concept in super-resolution imaging called MINFLUX (Balzarotti et al. 2017) and it can reach unprecedented resolutions for optical microscopy of ~6 nm so it is worth adding a brief commentary on how the work in this thesis might be related to this new concept. MINFLUX is based on combining two super-resolution imaging techniques together: STED and PALM/STORM, where a patterned excitation is used on a tiny field of view of the sample. If the pattern is a donut, a zero intensity center is being sequentially placed at four focal plane positions and photon counts are collected from each position. The concept revolves around the idea that if the fluorophore we are trying to localize is closer to one of the positions, when that position takes its turn in being at the central zero point there will be fewer photons collected from the fluorophore whereas for farther positions, more photons will be collected from the fluorophore. The collected information from the four positions can be used to statistically infer the position of the fluorophore with great accuracy and hence, the number of photons needed for localization is ~22-fold lower than the number required for other conventional centroid-localization methods.

This localization is then repeated for all fluorophores within the field of view by switching fluorophores on and off stochastically as with PALM/STORM.

Drift correction remains to be a big problem for MINFLUX as it takes tens of seconds to localize all fluorophores within the excitation region (~100 nm in diameter) where sample drift is expected to take place. Unfortunately, our drift correction algorithm won't be helpful for this method as the algorithm requires the detection of several molecules in a given frame to use their collective information for drift inference but MINFLUX setup allows for the detection of only one fluorophore per frame, which won't carry enough information for our algorithm. This is not just a problem for our method but it is a problem for image correlation as well. The use of fiducial markers is also problematic as their more stable on state might interfere with the localizations of the target fluorophores and also it is not guaranteed to find a fiducial marker within the desired tiny field of view. The remaining viable option that the authors used was to install a mechanical system lock in the microscopy setup that can automatically detect and correct for drift but even that solution wasn't very efficient and it added a significant complexity to the system, which made the authors admit that further improvement is required.

The localization precision of MINFLUX is significantly larger than other SMLM methods allowing for a ~6nm resolving power, however this higher precision comes at the price of having to do nanoscopy localizations within the ~100 nm field of view, which requires tens of seconds to acquire. This makes MINFLUX impractical in detecting fluorophore localizations in full super-resolution images of fixed samples and hence, the usefulness of the algorithms introduced in this thesis remains intact as conventional SMLM methods maintain their superiority over MINFLUX in acquiring full super-resolution images from fixed samples.

Overall, this work presents a significant advancement in the field of SMLM by providing cell biologists with innovative and efficient statistical tools that address the problems of sample drift and measurement noise.

References

Abassi, Yama a., and Kristiina Vuori. 2002. “Tyrosine 221 in Crk Regulates Adhesion-Dependent Membrane Localization of Crk and Rac and Activation of Rac Signaling.” *EMBO Journal* 21 (17): 4571–82. doi:10.1093/emboj/cdf446.

Abbe, E. 1873. “Beitr{ä}ge Zur Theorie Des Mikroskops Und Der Mikroskopischen Wahrnehmung.” *Archiv F{ü}r Mikroskopische Anatomie* 9 (1): 413–18. doi:10.1007/BF02956173.

Annibale, Paolo, Stefano Vanni, Marco Scarselli, Ursula Rothlisberger, and Aleksandra Radenovic. 2011. “Quantitative Photo Activated Localization Microscopy : Unraveling the Effects of Photoblinking.” *PloS One* 6 (7). doi:10.1371/journal.pone.0022678.

Asmussen, Søren, and Peter W. Glynn. 2011. “A New Proof of Convergence of MCMC via the Ergodic Theorem.” *Statistics and Probability Letters* 81 (10): 1482–85. doi:10.1016/j.spl.2011.05.004.

Balzarotti, Francisco, Yvan Eilers, Klaus C Gwosch, Arvid H Gynnå, Volker Westphal, Fernando D Stefani, Johan Elf, and Stefan W Hell. 2017. “With Minimal Photon Fluxes.” *Science* 355 (February): 606–12. doi:10.1126/science.aak9913.

Bar-On, Dana, Steve Wolter, Sebastian Van De Linde, Mike Heilemann, German Nudelman, Esther Nachliel, Menachem Gutman, Markus Sauer, and Uri Ashery. 2012. “Super-Resolution Imaging Reveals the Internal Architecture of Nano-Sized Syntaxin Clusters.” *Journal of Biological Chemistry* 287 (32): 27158–67. doi:10.1074/jbc.M112.353250.

Bates, Mark, Timothy R. Blosser, and Xiaowei Zhuang. 2005. “Short-Range Spectroscopic Ruler Based on a Single-Molecule Optical Switch.” *Physical Review Letters* 94 (10): 1–4. doi:10.1103/PhysRevLett.94.108101.

Bates, Mark, Bo Huang, Graham T. Dempsey, and Xiaowei Zhuang. 2007. “Multicolor Super-Resolution Imaging with Photo-Switchable Fluorescent Probes.” *Science* 317 (5845): 1749–53. doi:10.1126/science.1146598.

Baumgart, Florian, Andreas M Arnold, Konrad Leskovar, Kaj Staszek, Martin Fölser, Julian Weghuber, Hannes Stockinger, and Gerhard J Schütz. 2016. “Varying Label Density Allows Artifact-Free Analysis of Membrane-Protein Nanoclusters.” *Nature Methods* 13 (8). doi:10.1038/nmeth.3897.

Bayes, Mr., and Mr. Price. 1763. “An Essay towards Solving a Problem in the Doctrine of Chances. By the Late Rev. Mr. Bayes, F. R. S. Communicated by Mr. Price, in a Letter to John Canton, A. M. F. R. S.” *Philosophical Transactions of the Royal Society of London* 53 (0): 370–418. doi:10.1098/rstl.1763.0053.

Betzig, E., A. Lewis, A. Harootunian, M. Isaacson, and E. Kratschmer. 1986. “Near Field Scanning Optical Microscopy (NSOM).” *Biophysical Journal* 49 (1). Elsevier: 269–79. doi:10.1016/S0006-3495(86)83640-2.

Betzig, E., G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. 2006. “Imaging Intracellular Fluorescent Proteins at Nanometer Resolution.” *Science* 313: 1642–45.

Betzig, Eric, George H Patterson, Rachid Sougrat, O Wolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, and Harald F Hess. 2006. "Imaging Intracellular Fluorescent Proteins at Nanometer Resolution." *Science (New York, N.Y.)* 313 (5793): 1642–45. doi:10.1126/science.1127344.

Blitzstein, Joseph K., and Jessica Hwang. 2015. *Introduction to Probability*. doi:http://dx.doi.org/10.1016/S1363-4127(97)81322-2.

Chang, Kenneth. 2014. "Nobel Laureates Pushed Limits of Microscopes." *The New York Times*. https://www.nytimes.com/2014/10/09/science/nobel-prize-chemistry.html?_r=0.

Dempster, A. P., N. M. Laird, and D. B. Rubin. 1977. "Maximum Likelihood from Incomplete Data via the EM Algorithm." *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1): 1–38.

Donato, Dominique M, Larisa M Ryzhova, Leslie M Meenderink, Irina Kaverina, and Steven K Hanks. 2010. "Dynamics and Mechanism of p130Cas Localization to Focal Adhesions." *The Journal of Biological Chemistry* 285 (27): 20769–79. doi:10.1074/jbc.M109.091207.

Elmokadem, Ahmed, and Ji Yu. 2015. "Optimal Drift Correction for Superresolution Localization Microscopy with Bayesian Inference." *Biophysical Journal* 109 (9): 1772–80. doi:10.1016/j.bpj.2015.09.017.

Erni, Rolf, Marta D Rossell, Christian Kisielowski, and Ulrich Dahmen. 2009. "Atomic-Resolution Imaging with a Sub-50-Pm Electron Probe." *Phys. Rev. Lett.* 102 (9). American Physical Society: 96101. doi:10.1103/PhysRevLett.102.096101.

Fallis, A.G. 2013. "Moleculer Biology of the Cell." *Journal of Chemical Information and Modeling* 53 (9): 1689–99. doi:10.1017/CBO9781107415324.004.

Garcia-parajo, Maria F, Alessandra Cambi, Juan A Torreno-pina, and Nancy Thompson. 2014. "Nanoclustering as a Dominant Feature of Plasma Membrane Organization." *Journal of Cell Science*, 4995–5005. doi:10.1242/jcs.146340.

Geisler, Claudia, Thomas Hotz, Andreas Schönle, Stefan W. Hell, Axel Munk, and Alexander Egner. 2012. "Drift Estimation for Single Marker Switching Based Imaging Schemes." *Optics Express* 20 (7): 7274. doi:10.1364/OE.20.007274.

Geman, Stuart, and Donald Geman. 1984. "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images." *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-6 (6): 721–41. doi:10.1109/TPAMI.1984.4767596.

Giepmans, Ben N G. 2008. "Bridging Fluorescence Microscopy and Electron Microscopy." *Histochemistry and Cell Biology* 130 (2): 211–17. doi:10.1007/s00418-008-0460-5.

Glickman, Mark E, and David a van Dyk. 2007. "Basic Bayesian Methods." *Methods in Molecular Biology (Clifton, N.J.)* 404: 319–38. doi:10.1007/978-1-59745-530-5_16.

Gustafsson, Mats G L. 2005. "Nonlinear Structured-Illumination Microscopy: Wide-Field Fluorescence Imaging with Theoretically Unlimited Resolution." *Proceedings of the National Academy of Sciences of the United States of America* 102 (37): 13081–86. doi:10.1073/pnas.0406877102.

Hazan, Tamir, Francesco Orabona, Anand D. Sarwate, Subhransu Maji, and Tommi Jaakkola. 2016. "High Dimensional Inference with Random Maximum A-Posteriori Perturbations,"

1–44. http://arxiv.org/abs/1602.03571.

Heintzmann, Rainer, and Gabriella Ficz. 2013. “Breaking the Resolution Limit in Light Microscopy.” *Methods in Cell Biology* 114 (4): 525–44. doi:10.1016/B978-0-12-407761-4.00022-1.

Hell, Stefan W, Steffen J Sahl, Mark Bates, Romain F Laine, Gabriele S Kaminski Schierle, Sebastian Van De Linde, John R Allen, Stephen T Ross, and Michael W Davidson. 2015. “Single Molecule Localization Microscopy for Superresolution.” *Journal of Optics*. doi:10.1088/2040-8978/15/9/094001.

Hess, Samuel T, Thanu P K Girirajan, and Michael D Mason. 2006. “Ultra-High Resolution Imaging by Fluorescence Photoactivation Localization Microscopy.” *Biophysical Journal* 91 (11): 4258–72. doi:10.1529/biophysj.106.091116.

Huang, Bo, Mark Bates, and Xiaowei Zhuang. 2010. “Super Resolution Fluorescence Microscopy.” *Annual Review of Biochemistry* 78: 993–1016. doi:10.1146/annurev.biochem.77.061906.092014.Super.

Huang, Bo, Wenqin Wang, Mark Bates, and Xiaowei Zhuang. 2008. “Three-Dimensional Super-Resolution Imaging by Stochastic Optical Reconstruction Microscopy.” *Science (New York, N.Y.)* 319 (5864): 810–13. doi:10.1126/science.1153529.

Kiuchi, Tai, Makio Higuchi, Akihiro Takamura, Masahiro Maruoka, and Naoki Watanabe. 2015. “Multitarget Super-Resolution Microscopy with High-Density Labeling by Exchangeable Probes.” *Nature Methods* 12 (July). doi:10.1038/nmeth.3466.

Klar, T a, and S W Hell. 1999. “Subdiffraction Resolution in Far-Field Fluorescence Microscopy.” *Optics Letters* 24 (14): 954–56. doi:10.1364/OL.24.000954.

Lee, Sang Hak, Murat Baday, Marco Tjioe, Paul D. Simonson, Ruobing Zhang, En Cai, and Paul R. Selvin. 2012. “Using Fixed Fiduciary Markers for Stage Drift Correction.” *Optics Express* 20 (11): 12177–83.

Masters, Barry R. 2010. “The Development of Fluorescence Microscopy.” *Encyclopedia of Life Sciences*, 1–9. doi:10.1002/9780470015902.a0022093.

Matsuda, M., S. Tanaka, S. Nagata, A. Kojima, T. Kurata, and M. Shibuya. 1992. “Two Species of Human CRK cDNA Encode Proteins with Distinct Biological Activities.” *Molecular and Cellular Biology* 12 (8): 3482–89.

Mayer, B J, M Hamaguchi, and H Hanafusa. 1988. “A Novel Viral Oncogene with Structural Similarity to Phospholipase C.” *Nature* 332 (6161): 272–75. doi:10.1038/332272a0.

Mennella, V., B. Keszthelyi, K. L. McDonald, B. Chhun, F. Kan, G. C. Rogers, B. Huang, and D. A. Agard. 2012. “Subdiffraction-Resolution Fluorescence Microscopy Reveals a Domain of the Centrosome Critical for Pericentriolar Material Organization.” *Nature Cell Biology* 14 (11). Nature Publishing Group: 1159–68. doi:10.1038/ncb2597.

Mizuno, Kouichi, Takuji Fujita, Hiroshi Gomi, and Tetsuro Izumi. 2016. “Granuphilin Exclusively Mediates Functional Granule Docking to the Plasma Membrane.” *Nature Publishing Group*. Nature Publishing Group, 1–12. doi:10.1038/srep23909.

Mlodzianoski, Michael J., John M. Schreiner, Steven P. Callahan, Katarina Smolková, Andrea Dlasková, Jitka Šantorová, Petr Ježek, and Joerg Bewersdorf. 2011. “Sample Drift

Correction in 3D Fluorescence Photoactivation Localization Microscopy." *Optics Express* 19 (16): 15009. doi:10.1364/OE.19.015009.

Nagano, Makoto, Daisuke Hoshino, Naohiko Koshikawa, Toshifumi Akizawa, and Motoharu Seiki. 2012. "Turnover of Focal Adhesions and Cancer Cell Migration." *International Journal of Cell Biology* 2012 (January): 310616. doi:10.1155/2012/310616.

Patterson, G H, and Jennifer Lippincott-Schwartz. 2002. "A Photoactive GFP for Selective Photolabeling of Proteins and Cells." *Science* 297 (September): 1873–77. doi:10.1126/science.1074952.

Patterson, George, Michael Davidson, Suliana Manley, and Jennifer Lippincott-Schwartz. 2013. "Superresolution Imaging Using Single-Molecule Localizatio." *Amnu Rev Phys Chem*, no. 3: 345–67. doi:10.1146/annurev.physchem.012809.103444.Superresolution.

Rust, Michael J, Mark Bates, and Xiaowei W Zhuang. 2006. "Sub-Diffraction-Limit Imaging by Stochastic Optical Reconstruction Microscopy (STORM)." *Nat Methods* 3 (10): 793–95. doi:Doi 10.1038/Nmeth929.

Sage, Daniel, Hagai Kirshner, Thomas Pengo, Nico Stuurman, Junhong Min, Suliana Manley, and Michael Unser. 2015. "Quantitative Evaluation of Software Packages for Single-Molecule Localization Microscopy." *Nature Methods* 12 (8). doi:10.1038/nmeth.3442.

Sawada, Yasuhiro, Masako Tamada, Benjamin J Dubin-Thaler, Oksana Cherniavskaya, Ryuichi Sakai, Sakae Tanaka, and Michael P Sheetz. 2006. "Force Sensing by Mechanical Extension of the Src Family Kinase Substrate p130Cas." *Cell* 127 (5): 1015–26. doi:10.1016/j.cell.2006.09.044.

Schermelleh, Lothar, Rainer Heintzmann, and Heinrich Leonhardt. 2010. "A Guide to Super-Resolution Fluorescence Microscopy." *Journal of Cell Biology* 190 (2): 165–75. doi:10.1083/jcb.201002018.

Sharonov, Alexey, and Robin M Hochstrasser. 2006. "Wide-Field Subdiffraction Imaging by Accumulated Binding of Diffusing Probes." *Proceedings of the National Academy of Sciences of the United States of America* 103 (50): 18911–16. doi:10.1073/pnas.0609643104.

Shroff, H, C G Galbraith, J A Galbraith, and E Betzig. 2008. "Live-Cell Photoactivated Localization Microscopy of Nanoscale Adhesion Dynamics." *Nature Methods* 5 (5): 417–23. doi:10.1038/nmeth.1202.

Small, Alexander R, and Raghuveer Parthasarathy. 2014. "Superresolution Localization Methods." *Annual Review of Physical Chemistry* 65: 107–25. doi:10.1146/annurev-physchem-040513-103735.

Smith, Alex J., and Alan S. Verkman. 2015. "Superresolution Imaging of Aquaporin-4 Cluster Size in Antibody-Stained Paraffin Brain Sections." *Biophysical Journal* 109 (12). Biophysical Society: 2511–22. doi:10.1016/j.bpj.2015.10.047.

Sonnen, Katharina F., Lothar Schermelleh, Heinrich Leonhardt, and Erich A. Nigg. 2012. "3D-Structured Illumination Microscopy Provides Novel Insight into Architecture of Human Centrosomes." *Biology Open* 1 (10): 965–76. doi:10.1242/bio.20122337.

Tang, Y., X. Wang, X. Zhang, J. Li, and L. Dai. 2014. "Sub-Nanometer Drift Correction for Super-Resolution Imaging." *Optics Letters* 39 (19): 5685–88.

Tatavarty, Vedakumar, Eun Ji Kim, Vladimir Rodionov, and Ji Yu. 2009. “Investigating Sub-Spine Actin Dynamics in Rat Hippocampal Neurons with Super-Resolution Optical Imaging.” *PLoS ONE* 4 (11). doi:10.1371/journal.pone.0007724.

Thompson, Russell E, Daniel R Larson, and Watt W Webb. 2002. “Precise Nanometer Localization Analysis for Individual Fluorescent Probes.” *Biophysical Journal* 82 (5): 2775–83. doi:10.1016/S0006-3495(02)75618-X.

Veatch, Sarah L., Benjamin B. Machta, Sarah A. Shelby, Ethan N. Chiang, David A. Holowka, and Barbara A. Baird. 2012. “Correlation Functions Quantify Super-Resolution Images and Estimate Apparent Clustering due to over-Counting.” *PLoS ONE* 7 (2). doi:10.1371/journal.pone.0031457.

Vogelsang, Jan, Robert Kasper, Christian Steinhauer, Britta Person, Mike Heilemann, Markus Sauer, and Philip Tinnefeld. 2008. “A Reducing and Oxidizing System Minimizes Photobleaching and Blinking of Fluorescent Dyes.” *Angewandte Chemie - International Edition* 47 (29): 5465–69. doi:10.1002/anie.200801518.

Wang, Yina, Joerg Schnitzbauer, Zhe Hu, Xueming Li, Yifan Cheng, Zhen-Li Huang, and Bo Huang. 2014. “Localization Events-Based Sample Drift Correction for Localization Microscopy with Redundant Cross-Correlation Algorithm.” *Optics Express* 22 (13): 15982. doi:10.1364/OE.22.015982.

Waters, Jennifer C. 2009. “Accuracy and Precision in Quantitative Fluorescence Microscopy.” *The Journal of Cell Biology* 185 (7): 1135–48. doi:10.1083/jcb.200903097.

Wegel, Eva, Antonia Göhler, B. Christoffer Lagerholm, Alan Wainman, Stephan Uphoff, Rainer Kaufmann, and Ian M. Dobbie. 2016. “Imaging Cellular Structures in Super-Resolution with SIM, STED and Localisation Microscopy: A Practical Comparison.” *Scientific Reports* 6 (1). Nature Publishing Group: 27290. doi:10.1038/srep27290.

Yildiz, A, J N Forkey, S A McKinney, T Ha, Y E Goldman, and P R Selvin. 2003. “Myosin V Walks Hand-over-Hand: Single Fluorophore Imaging with 1.5-Nm Localization.” *Science* 300 (5628): 2061–65. doi:DOI 10.1126/science.1084398.

Zaidel-Bar, R, M Cohen, L Addadi, and B Geiger. 2004. “Hierarchical Assembly of Cell-Matrix Adhesion Complexes.” *Biochemical Society Transactions* 32 (Pt3): 416–20. doi:10.1042/BST0320416.

Zheng, J., K. Machida, S. Antoku, K. Y. Ng, K. P. Claffey, and B. J. Mayer. 2010. “Proteins That Bind the Src Homology 3 Domain of CrkI Have Distinct Roles in Crk Transformation.” *Oncogene* 29 (48): 6378–89. doi:10.1038/onc.2010.369.