Feedback Tracking and Correlation Spectroscopy of Fluorescent Nanoparticles and Biomolecules

Thesis by

Kevin L. McHale

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2008
(Defended January 8, 2008)

Acknowledgments

Work toward this thesis began over twenty years ago when my parents, Beth and Michael, set me on the course that led to it. They will freely admit that for some time now their guidance has tended more toward the personal than the academic, but they have taught me to be the person that I am today and I can think of nothing to be more thankful for than that. I thank my sisters, Sara and Marissa, with whom growing up was always exciting and being grown up is always rewarding. Thank you to my grandmother Bernice and my extended family, in particular my cousins Michael and Linda Pisano. I also thank my newest family members, Robin, Ivy and Ian Hest. This thesis would not have been possible without all of my family's support.

Hideo Mabuchi has been all that I could have asked for in an advisor. His patience, creativity, and confidence in the abilities of his students allows him to run a lab like no other I am aware of. The freedom afforded to me to explore my own ideas and solve problems (or to try, at least) in my own way contributed tremendously both to my learning and to my overall happiness. He has taught me how to be a scientist, and provided a model for leadership that I can only hope to emulate.

I am thankful for the guidance I have received over the years from several other scientists at Caltech. I enjoyed a rotation with Demetri Psaltis during my first year, and I appreciate and admire his willingness to spare as much time as he did. I thank my candidacy and thesis committee members, Niles Pierce, Erik Winfree, John Doyle, Changhuei Yang and Zhen-Gang Wang for their time, interest, insight and suggestions. I thank Shimon Weiss at UCLA for being a friendly link to the single-molecule community and for helping me put my work into perspective. I am also thankful to Wallace Brey, my undergraduate advisor, and Bethany Bechtel for introducing me to scientific research. Their guidance in my very early years as a scientist was an essential factor in my decision to attend graduate school.

While Hideo taught me to be a scientist, my fellow members of Hideo's group taught me to actually *do* science and became good friends in the process. I was very fortunate to work closely with Andy Berglund during my first few years in the lab. The amount that I learned from him, and in later years with him, accounts for a large proportion of that which I learned in grad school. I am grateful to Mike Armen, who never once complained about my unrelenting use of

him as a sounding board for (often far-fetched) ideas. I thank John Stockton *despite* not having written that basketball paper, and I thank Tony Miller for, well, keeping the group keen. I thank John Au for his foosball moves, Tim McGarvey for telling me about the world, Nathan Hodas for falling down a mountain with me and Luc Bouten for perpetual encouragement, friendly competition and unsolicited investment advice. Special thanks go to Sheri Stoll, my daily source of good-natured argument and the principal reason our lab ran so smoothly. Thanks also to Paul Barclay, Felipe Caycedo-Soler, André Conjusteau, Orion Crisafulli, Nicole Czakon, JM Geremia, Asa Hopkins, Joe Kerckhoff, Kenny Kwon, Ben Lev, Gopal Sarma, Andrew Silberfarb, Jen Sokol, Ramon van Handel and Jon Williams. I enjoyed working with all of you. Finally, I wish the best of luck to Charles Limouse and Michael Zhang in carrying on where I leave off.

I was fortunate to enter a Bioengineering department filled with great people. My classmates Jennifer Tom, Gwyneth Card, Jeff Endelman, Frank Lee, Derek Rinderknecht, Anna Grosberg and Princess Imoukhuede helped keep our first year fun despite certain obstacles. I shared many good times with Ben Lin, Bodhi Sansom and John Dabiri, and I will never forget the late-night trek from Ventura to Pasadena that Mike Reiser so kindly accompanied me on. I will forever be proud of founding the BEars softball team, and I am glad to leave it in the hands of Emily McDowell.

Many of my favorite times at Caltech were spent with my roommate of four years, Joe Zadeh. The story of our adventures could fill a volume at least the size of this thesis. I found great friends in Jane Khudyakov, Alexa Price-Whelan and Joe Schramm, and we had many fun times as well. I enjoyed my year playing on the Caltech rugby team and the times I shared with the team both on and off the pitch; my frequent visits to Brad Cenko's office; dinner parties with Talia Starkey; hijinks with any combination of Will Ford, Matt Lucas and Jason Keith; dragon boat racing; foosball, Mügen, croquet and, of course, Del Taco. Most of all, I am grateful for the companionship of Megan Elizabeth Eckart, my greatest advocate and closest friend, whose support is the only thing that got me through my tumultuous final month as a grad student at Caltech.

I am fortunate to have a number of friends from home who, despite the obstacles of thousands of miles, three time zones and busy lives, remain close to me to this day. While keeping in touch is not always easy, it is a true testament to my friendships with David Abramson, Samantha Martin, Steve Kay, Phillip Crandall and Mike Flamino that while our lives are always changing, things between us always seem to stay the same. All of you are on my mind more than you know.

Finally, I thank the Charles Lee Powell Foundation for financial support during my first year at Caltech, and the National Defense Science and Engineering Graduate Fellowship for support during years three through five.

Abstract

The best way to study dynamic fluctuations in single molecules or nanoparticles is to look at only one particle at a time, and to look for as long as possible. Brownian motion makes this difficult, as molecules move along random trajectories that carry them out of any fixed field of view. We developed an instrument that tracks the Brownian motion of single fluorescent molecules in three dimensions and in real-time while measuring fluorescence with nanosecond time resolution and single-photon sensitivity. The apparatus increases observation times by approximately three orders of magnitude while improving data-collecting efficiency by locking tracked objects to a high-intensity region of the excitation laser.

As a first application of our technique, we tracked and studied the fluorescence statistics of semiconductor quantum dots. Our measurements were well resolved at 10ns correlation times, allowing measurement of photon anti-bunching on single particles in solution for the first time. We observed variations of (34 ± 16)% in the fluorescence lifetimes and (23 ± 18)% in the absorption cross-sections within an aqueous quantum dot sample, confirming that these variations are real, not artifacts of the immobilization methods previously used to study them. Additionally, we studied quantum dot fluorescence intermittency and its dependence on 2-mercaptoethanol, finding evidence that the chemical suppresses blinking on short time-scales (< 1s) by reducing the lifetime of the dark state.

Finally, we studied the translational and intramolecular Brownian motion of λ-phage DNA molecules. Our apparatus decouples these motions almost completely, and yielded a translational diffusion coefficient estimate $D = (0.71 \pm 0.05)\mu\text{m}^2/\text{s}$ lying between previous measurements for this molecule under identical solution conditions but with less precise techniques. Our measurements show clear evidence of intramolecular motion of the polymer chain in the form of statistical correlations on time-scales up to 1s, but we have not yet been able to determine the influence of solvent interactions on these dynamics.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

We are only beginning to understand the degree to which biology is influenced by consequences of the size of its basic building blocks. Nucleic acids and proteins are special in that they are big, but not too big, relative to the size of the water molecules that surround them. Under constant bombardment by surrounding fluid, biological molecules bend, stretch, vibrate, fold and unfold, break apart and recombine; these dynamics almost certainly influence the way in which they interact with each other. Small molecules do not behave like this because the stiffness of their chemical bonds prevents such structural fluctuations; nor do large objects, because viscous drag quickly dampens any acceleration due to small collisional forces. These types of motions are uniquely the domain of large molecules, most often biological molecules, and their exact characteristics and the extent of their consequences are still widely unknown.

Dynamic fluctuations in the shapes of molecules are hard to detect because no two molecules experience the same sequence of collisions, so no two molecules exhibit the same fluctuations. Any attempt to study such fluctuations among large numbers of molecules will fail, because the motion of each molecule is drowned out by the motion of all the others. For most of the history of science, we have therefore been limited to the measurement of the average, equilibrium properties of microscopic biological systems.

The first successful measurement of nonequilibrium molecular dynamics occurred in 1972[1]. Incredible technological development enabled the detection of single fluorescent molecules in solution in 1990[2] and single-molecule detection has since exploded in popularity, with new applications and techniques under constant development. Single-molecule measurements have revealed, for example, dynamic variations in the shapes and orientations of proteins, both spontaneously and as a result of interactions with small molecules[3]; directed motion of small “motor” proteins along larger ones[4, 5]; nanosecond-scale reconfiguration dynamics in unfolded proteins[6]; and proteins that undergo entire folding transitions on microsecond time-scales[7].

The most popular methods for single-molecule detection are based on laser-induced fluorescence; a basic schematic of a typical experiment is shown in Fig. 1.1. A laser beam is

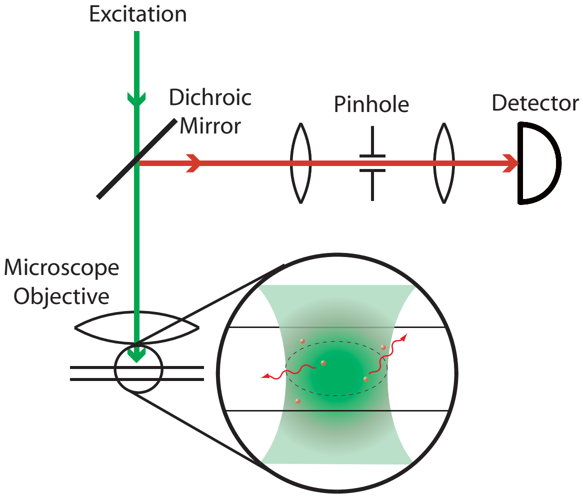


Figure 1.1: Example single-molecule fluorescence experiment. The pinhole reduces background from out-of-focus fluorescence emitters.

focused by a high power microscope objective to a small, intense spot within a liquid sample. Fluorescent molecules — or molecules labeled by a fluorescent probe — drift into and out of the beam as they move by Brownian motion, the translational motion due to collisions with solvent molecules. The fluorescence emitted by the molecules is collected by the objective and focused onto a sensitive detector. The resulting signal is either a high time-resolution sequence of fluorescence bursts, like that illustrated in Fig. 1.2, or a lower time-resolution but spatially resolved image of the molecule and its surroundings.

Single-molecule fluorescence measurements have some limitations that are apparent just from the appearance of the data in Fig. 1.2. The duration of each burst is determined by the molecule's motion, and typically limited to less than 10ms (longer for imaging methods, but still limited). This places restrictions on the time-scales that can be probed using these methods. Many biological molecules have interesting dynamics that are on time-scales as long as or longer than this, and therefore that cannot be resolved by these measurements. In addition, we only collect a finite number of photons from each molecule, and this number determines how well the molecule's dynamics can be resolved: the fewer pairs of photons are detected with time-spacing Δt, the larger the measurement uncertainty becomes on that time-scale. This fundamental type of noise arises any time a process is described by a continuous rate but mediated by discrete events[8, e.g.]. We could try to reduce this noise by increasing the intensity of the excitation laser, but dye saturation and bleaching will always limit our ability to do this.

In order to make high time-resolution or long-duration measurements on a single molecule, we must look at it for a long period of time. The simplest way to do this is by attaching it to something that prevents it from moving out of the focus of the microscope. However, if our goal

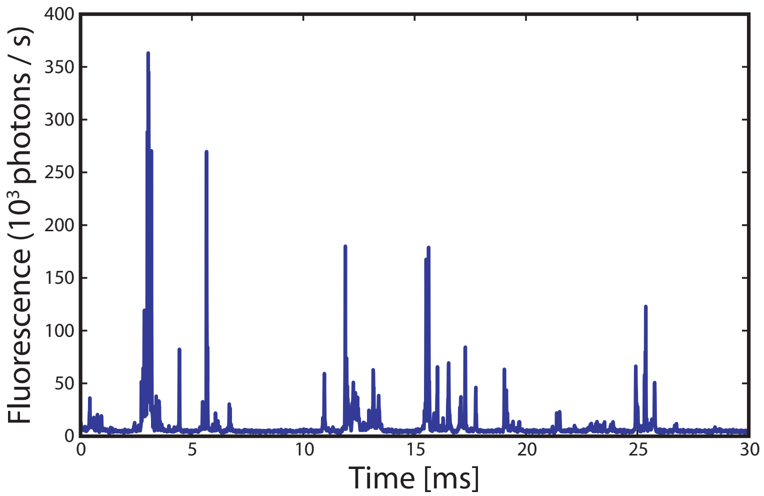


Figure 1.2: Example single-molecule fluorescence signal. Molecules emit bursts as they drift through the focus of the excitation laser.

is to understand biology then we should be concerned about the influence of the attachment surface on the molecule stuck to it. At the very least we expect it to perturb the molecule's structure by concentrating at the point of attachment all of the forces incident on the molecule due to solvent collisions. The chemistry of the attachment surface can influence the molecule as well: glass surfaces, for example, often carry an electrical charge that will affect the molecule through electrostatic forces.

Over the last few years, several groups[9–15] have begun to develop a new technology for extending single-molecule observation times using feedback control. These approaches allow the molecule to move freely in solution while active hardware moves some combination of the microscope optics or sample stage in order to keep the molecule in focus at all times. These techniques in principle would allow for the completely passive observation of individual molecules often for times that are orders of magnitude longer than they would be in the absence of feedback. This thesis documents my work, primarily over the past 18 months, in developing what is arguably the most capable of these techniques yet developed.

1.1 Personal history

My path to Caltech resembled the random trajectories of the molecules described in this thesis. As an undergraduate I had difficulty focusing my ambition, and this resulted in having two major subjects, a minor subject, complete pre-medical preparation and in applying to four different types of graduate programs. I came to Caltech less than two months after deciding that I did not want to be an academic physician, and I did not have any concrete plans for what

I would do once I arrived.

I knew when I committed to Caltech that I would be a part of a very young Bioengineering program, but I failed to fully anticipate the extent to which this would affect me. Fortunately I had a good group of friends in my classmates, and by supporting each other we made it through the rough times and — I hope — helped improve the program for those that succeeded us.

I began doing research in my second term, when I worked with Demetri Psaltis on a project comparing the imaging properties of insect-inspired compound lens systems with more traditional simple lens systems. My goal in this project was to determine applications for which a synthetic version of such an imaging system might be particularly useful. Over the course of the term I — with the help of insightful discussions with Michael Dickinson at Caltech and Simon Laughlin at the University of Cambridge — came to the conclusion that compound eyes did not evolve for their imaging properties, and that a compound lens system would likely only be favored in applications in which imaging performance was not the primary design parameter. Demetri's guidance on this project kept me excited about doing research, and the project itself sparked an early interest in optics that would develop through the remainder of my time in grad school.

I attribute my meeting of Hideo Mabuchi to serendipity alone. In my first term at Caltech I registered for a course in control theory, a subject I had never previously studied. One component of that course was a weekly guest lecture on applications of control theory in Caltech labs. I did not attend those lectures regularly, but happened to attend one that Hideo gave about laser stabilization, Michael Armen's adaptive homodyne experiment and the early stages of John Stockton's magnetometry experiment. I thought it was interesting, and a few months later I found myself emailing Hideo about doing a research rotation in his lab.

Hideo answered my email almost immediately, we met, and he suggested I talk to Andrew Berglund about some of the recursive estimation algorithms he had been working on for his FRET microscopy experiment. I began working with Andy on a statistical method for identifying molecules based on individual fluorescence bursts in single-molecule experiments, and over the course of that term I wrote code to simulate fluorescence data and implement the identification algorithm. I officially joined Hideo's lab over the summer, and we submitted a paper on my work in December 2003[16].

Over the next year, I bounced between a few different projects both in and out of the lab. Early on I taught myself to build basic feedback electronics, with the goal of temperature-stabilizing Andy's microscope. The stage that I built worked fairly well, but we never ended up using it because the timing did not work out: I had more theoretical projects I was interested in, and Andy began to redesign the apparatus for particle tracking experiments. I spent some time studying diffusion-limited chemical kinetics, focusing in particular on ideas born from discussions with Hideo, Andy and Igor Mezic at UCSB about the relationship between mixing and reaction statistics in extremely dilute solutions. This work, combined with some comments from a referee of my first paper, motivated me to write a paper about the effects of counting noise on estimates of the concentration of dilute solutions.

While working on concentration estimation, I took a bit of a detour to work with John Au on his application of convex optimization techniques to the design of photonic bandgap crystal structures. I had studied convex optimization during my first year at Caltech, and I was glad to be able to revisit the subject with a real application to work on. For a few months, John and I worked in parallel: he developed the physical theory and I wrote optimization software that was custom-tailored to his problem. In the end my code became a fairly robust interior-point solver, we both learned a lot about optimization, and John went on to develop an iterative method for designing photonic crystal structures with optical properties that are potentially useful for atomic physics experiments.

I finished my candidacy exam in November 2005, proposing to apply my identification and concentration estimation algorithms to problems in molecular biophysics. Almost immediately afterward, I began working on something very different and never turned back. As an apology to my candidacy committee, this was truly not my intention — I distinctly remember a conversation with Megan Eckart about the fluorescence microscopy experiment in which she asked why we needed so many optics and I replied that those were for the tracking apparatus, and that I “had no interest” in that. Andy and I began working together under the premise that he would teach me to use his microscope so that I could work on my proposed experiments while he was working on the tracking apparatus. Literally within hours we had taken the entire apparatus apart, and so began my work on particle tracking.

1.2 Experimental history

The apparatus described in this thesis is the third (or fourth, depending on what counts) generation of the fluorescence microscopy experiment in Hideo Mabuchi’s lab at Caltech (now Stanford University). The original apparatus was a standard single-molecule fluorescence microscope built by Andrew Berglund to study fast dynamics of resonant energy transfer between fluorescent dyes[17]. Motivated by a closed-loop tracking proposal by Jörg Enderlein[9, 10], Andy began modifying his microscope to enable feedback tracking of fluorescent particles late in 2003[12]. Over the years that followed, he developed a theoretical framework for describing the statistics and limitations of feedback particle tracking systems and demonstrated his first experimental success in a paper in 2005[18], in which he tracked 60nm fluorescent spheres in solution and accurately characterized the tracking statistics that he measured.

I began working on the apparatus with Andy in late 2005 and we quickly realized that it needed to be rebuilt as a dedicated tracking microscope in order to optimize its performance. Over the next 6 months, through improvements of the mechanical and electronic components, beam and detector alignment, background noise rejection, beam intensity control and general experimental technique, we developed an apparatus that was capable of tracking 60nm and 210nm fluorescent nanospheres in two dimensions with accuracy limited almost entirely by photon-counting noise (with a signal size of about 4500 photons/sec)[19]. With our experiments complete, Andy began writing his thesis while I set out to completely rebuild the apparatus once again.

I began constructing an apparatus for tracking single fluorescent particles in three dimensions during the summer of 2006, basing it on Hideo's ideas for using a pair of laser beams to track particles along the optic axis of the microscope. That apparatus began operating in November 2006, and with it I demonstrated tracking of individual semiconductor quantum dots freely-diffusing in water[20]. I was able to resolve fluorescence anti-bunching on single quantum dots in solution, allowing me to characterize heterogeneity in quantum dot fluorescence statistics as well as fluorescence intermittency on time-scales as long as 1s. These measurements highlighted the importance of tracking techniques for making new types of measurements on single fluorescent objects, and also represented a technological advance in that no comparable technique had yet tracked objects that moved this quickly.

More recently, I have focused on tracking single DNA molecules for the evaluation of theoretical predictions for the internal Brownian dynamics of polymers in solution. While several measurements of this type have been made over the past few years[21–25], it is my view that none combines sufficient theoretical rigor with experimental data that can provide conclusive arguments for or against particular predictions. This is the motivation for my work in this area.

1.3 Organization of the thesis

I have made an effort to make this thesis entirely self-contained, which applies primarily to its theoretical components. In the first two chapters, this means including detailed reviews of relevant work in particle-tracking experiments before introducing my own work that is either specifically relevant to tracking in three dimensions or just an aspect of particle tracking dynamics that I find interesting. The last two chapters deal with applications of the tracking apparatus that themselves require a bit of theory, and those theories are reviewed with a level of detail dictated by the applications.

Chapter 2 provides the basic ideas behind the most important part of the apparatus — the method for sensing the position of fluorescent particles in three dimensions. The chapter begins with a review of the theory behind two-dimensional localization, most of which was developed by Andy Berglund. The chapter continues by extending the discussion to three-dimensional localization, in which I derive the basic properties of the localization method that we developed.

Chapter 3 reviews the dynamics of the tracking system. I discuss the feedback loop that the tracking system is based on, and review the linearized theory developed by Andy Berglund for describing the statistics of the tracking system in a "good" tracking limit. I conclude the chapter by deriving some of consequences of position estimation nonlinearity on the tracking statistics — in particular, I find nearly exact expressions for the statistics of particle escape from a tracking system by using a first-passage time approach.

Chapter 4 begins with a basic review of Fluorescence Correlation Spectroscopy (FCS)[1, 26], a single-molecule technique that we apply to our measured fluorescence signals. Next, I derive in a very general way the statistics of FCS for a tracked particle with internal degrees of translational freedom in preparation for the study of the intramolecular dynamics of DNA molecules.

Chapter 5 describes the experimental apparatus in detail.

Chapter 6 discusses the application of the tracking apparatus to the study of semiconductor quantum dots. The chapter begins with a background discussion on quantum dots, and develops a theoretical model for describing the basic properties of their fluorescence statistics. Next I present my measurements of photon anti-bunching in freely-diffusing quantum dots, and a detailed analysis of those measurements. I conclude with a basic theoretical description of quantum dot blinking statistics, and present our measurements in this area.

Chapter 7 describes the application of the apparatus to the study of the dynamics of the motion of λ-phage DNA. This chapter begins with a detailed discussion of the basic polymer dynamics theories that are necessary for understanding some of the arguments, mostly against the validity of certain approaches taken in the literature, later in the chapter. It concludes with a set of measurements on both the translational and intramolecular dynamics of the DNA molecules. This work remains ongoing.

Chapter 8 is a brief conclusion and discussion.

I include the two papers that I wrote before starting work on the tracking experiments as Appendices A and B. The former describes an approach by which the fluorescence signal from single-molecule experiments is analyzed in the time domain using a recursive Bayesian estimator, and was published in 2004[16]. The latter is a detailed study of the statistics of concentration fluctuations in the context of concentration measurement via single-molecule detection. In it I derive an information bound on the uncertainty in concentration estimates that is primarily applicable to single-molecule detection experiments, though holds true for all concentration measurements. Next I develop a time-domain concentration estimation technique based on the statistics of the waiting times between single-molecule detections, and evaluate its performance on simulated data and in comparison to alternative methods. It will be published eventually, but likely in a less-detailed form.

1.4 Publications resulting from graduate work

While at Caltech my work resulted, at least in part, in several publications that have either appeared in peer-reviewed journals or are almost ready to be submitted.

1. K. McHale, A. J. Berglund and H. Mabuchi. "Bayesian estimation for species identification in single-molecule fluorescence spectroscopy." *Biophys. J.* **86**, 3409 (2004).
2. K. McHale, A. J. Berglund and H. Mabuchi. "Near-optimal dilute concentration estimation via single-molecule detection." *In preparation.*
3. J. K. Au, K. McHale and H. Mabuchi. "Inverse problem based design of photonic crystal devices — Part II." *In preparation.*
4. A. J. Berglund, K. McHale and H. Mabuchi. "Feedback localization of fluorescent particles near the optical shot-noise limit." *Opt. Lett.* **32**, 145 (2007).
5. A. J. Berglund, K. McHale and H. Mabuchi. "Fluctuations in closed-loop fluorescent particle tracking." *Opt. Express* **15**, 7752 (2007).
6. K. McHale, A. J. Berglund and H. Mabuchi. "Quantum dot photon statistics measured by three-dimensional particle tracking." *Nano Lett.* **7**, 3535 (2007).

In addition, we filed for a provisional United States patent for the three-dimensional tracking apparatus.

Chapter 2

Three-dimensional localization of a fluorescent particle

Both the most difficult and the most important task in closed-loop particle tracking is the estimation of the position of the particle being tracked. Tracking errors must be sensed accurately and quickly in order for the feedback system to correct them before the particle escapes. Any noise intrinsic to the localization method will limit its accuracy; any latency in the localization estimate will limit the tracking bandwidth. Furthermore, the sensitivity of the localization method will often determine the amount of time for which particles can be tracked: the more photons are needed for localization, the faster the tracked particle bleaches and the sooner it becomes invisible to the apparatus.

We begin this chapter with a discussion of the symmetry-breaking requirements necessary for measuring three-dimensional position estimates for a single fluorescent particle, and we discuss several methods that involve symmetry-breaking in the fluorescence detector. We next introduce temporal symmetry breaking using spatial beam modulation, discussing the method used by several groups, including ours, for two-dimensional localization and tracking. Finally, we present the method that we have developed for three-dimensional localization and conclude with a derivation of the localization noise statistics intrinsic to this method.

2.1 Beam symmetry and DC symmetry breaking

Traditional single-molecule fluorescence microscopes, such as that depicted in Fig. 1.1, employ non-imaging detectors that do not provide enough information to determine the positions of detected particles because both the excitation laser beam and the collection and detection optics are highly symmetric. A focused TEM00 Gaussian laser beam has the intensity profile[27]

$$I(x, y, z) = \frac{2P}{\pi w^2(z)} \exp\left[-2\frac{x^2 + y^2}{w^2(z)}\right], \tag{2.1}$$

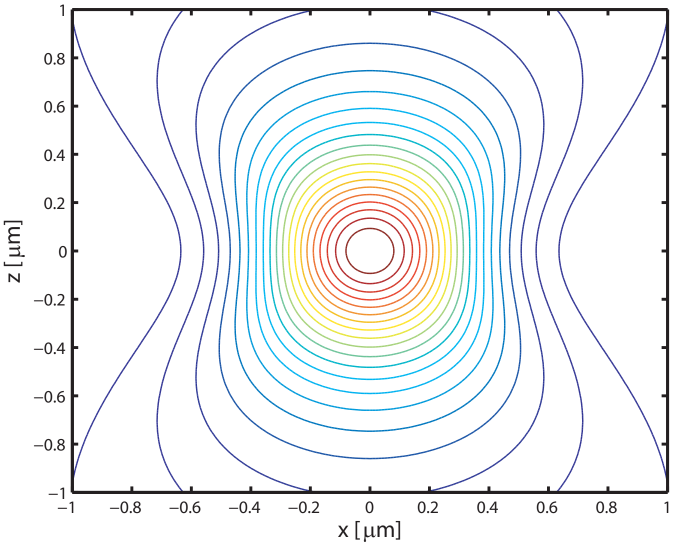


Figure 2.1: Contour plot of a diffraction-limited 532nm Gaussian beam.

where P is the total optical power in the beam, z is the axis of beam propagation, the $z-$dependent waist is given by

$$w^2(z) = w_0^2 \left[1 + \left(\frac{z\lambda}{\pi w_0^2} \right)^2 \right], \tag{2.2}$$

and we have allowed the beam centroid to lie at the coordinate origin. We sometimes define the Rayleigh range $z_R = \pi w_0^2/\lambda$ for notational simplicity in later sections. Figure 2.1 shows a cross-section of this profile for a beam with $\lambda = 532$nm that is focused to its diffraction limit ($w_0 = \lambda/2$). The consequences of the beam symmetry, as it relates to tracking, are obvious from the contour plot: because the only measurement available to us is the fluorescence rate, we cannot distinguish between any points that lie on the same contour surface. In order to track, something must be done to break this symmetry.

Symmetry-breaking approaches tend to fall into two categories: either the symmetry of the beam is broken while the detection optics are unchanged, or vice-versa. Outside of our group, most methods rely on breaking symmetry in the detection optics using imaging detectors, such as quadrant photodiodes or charge-coupled device (CCD) cameras. These detectors have good two-dimensional spatial resolution (along the x and y axes) and have been used for localization in several closed-loop particle tracking experiments[13–15, 28–31]. However, no optical detector technologies offer three-dimensional position sensitivity, so additional work is necessary to localize particles in three dimensions.

One way to determine the z position of a tracked particle is to use a cylindrical lens to create an optical aberration in which the fluorescence from a point emitter creates an elliptical image, and the relationship between the sizes of the ellipse axes is determined by the particle's displacement from the focal plane of the laser beam[32]. A CCD records images of the fluorescence, and computer software computes position estimates in real time based on the shape of the fluorescence ellipse. A second method tracks the particle along the x and y axes using a quadrant detector and locks the particle's z position in a plane either above or below the xy plane. The combination of the laser beam's divergence and the decaying detection efficiency along the z axis induced by the use of a confocal pinhole creates a steep enough intensity gradient to estimate the particle's position directly from the detected fluorescence intensity: if the intensity is too high (low), the tracking system should move the particle further from (closer to) the beam centroid[15, 31]. These approaches have been used to track fluorescent particles in viscous solutions, with diffusion coefficients as high as $0.6\mu\text{m}^2/\text{s}$, roughly corresponding to a 700nm diameter sphere in water.

The localization techniques described in this section all encode position information in low-frequency components of the fluorescence — either as the size of the differential signal between pixels in an imaging detector or simply as the absolute fluorescence intensity. Such position estimates are subject to systematic low-frequency noise sources that limit their tracking fidelity. The sensitive photodetectors necessary for detecting single molecules will always detect a background signal from scattering and auto-fluorescence in the glass coverslides, ambient scattered laser light, room lighting in the laboratory, power LEDs on equipment, etc. Any fluctuations in the size of these background signals — due to, for example, spatially nonuniform impurities in coverslides or 60Hz noise from AC power lines — feed into the position estimates and cause tracking errors. Even more importantly, any internal dynamics of the particles being tracked — blinking of dyes or quantum dots, or conformational motion of a polymer, for example — also couple into the localization estimates. Imaging methods tend to be less sensitive to these types of noise than intensity-based methods because they are *balanced*: such noise is correlated across all pixels on the detector, but only differential signals appear in the position estimates. This may explain why two-dimensional tracking methods using imaging detectors have greatly outperformed three-dimensional methods that incorporate intensity-based z localization[30, 31]. However, no detector is perfectly balanced and even the detectors themselves, in the absence of incident light, produce low-frequency noise that can have a significant impact on the localization signals. For more information on the ubiquity of low-frequency noise, see [33, 34].

A final commonality among the localization methods described in this section is the high cost and complexity involved in their implementation. An electron-multiplying CCD camera can cost as much as $\$40,000$, and five avalanche photodiode modules, each costing $8000, were needed for tracking based on an axial intensity gradient[31]. In the next section we transition to

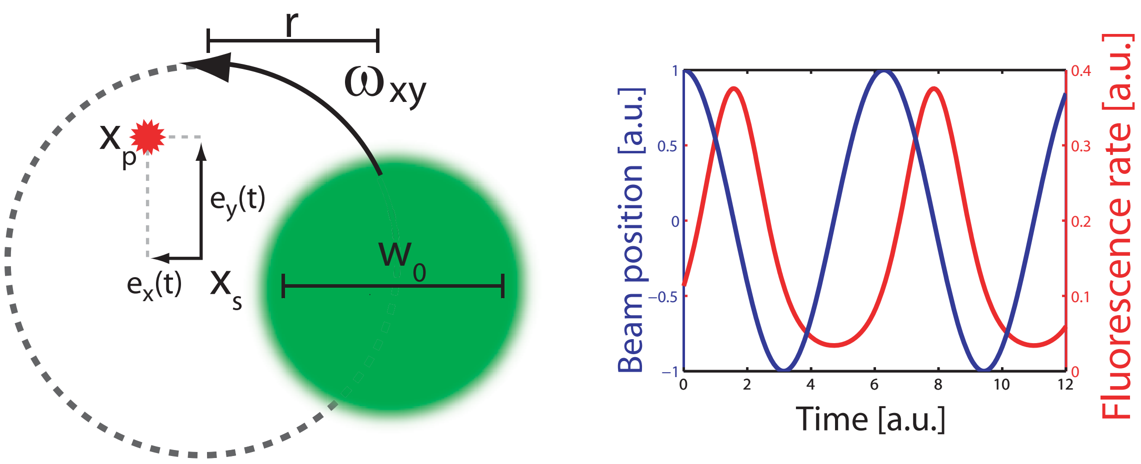


Figure 2.2: Beam rotation and resulting frequency modulation of the fluorescence intensity.

breaking the symmetry of the excitation beam while using a very simple detector arrangement. This method is both insensitive to low-frequency noise and much less expensive.

2.2 Particle localization in two dimensions via spatial beam modulation

In 2000, Jörg Enderlein proposed using a laser beam, deflected rapidly in a circular orbit, for the two-dimensional localization of single fluorescent particles[9, 10]. The idea is actually quite simple: the fluorescence is brightest whenever the beam is closest to the particle, and dimmest whenever the beam is furthest away. The resulting fluorescence intensity varies sinusoidally at the rotation frequency; its amplitude is proportional to the particle's distance from the rotation axis, and its phase relative to the circular deflection is precisely the angle corresponding to the particle's position in polar coordinates. By determining the magnitude and phase of the oscillating fluorescence signal, we can derive two-dimensional position estimates. Figure 2.2 illustrates this approach.

The first experimental realization of this idea was made by Enrico Gratton's group[11, 35], and Andy Berglund in our group followed suit shortly afterward[18]. Andy followed his experimental work with theoretical work describing localization noise and tracking limits [36]. This section serves as a simplified review of that work, in preparation for our discussion in the next section of the three-dimensional localization method that we have developed. The calculations in that section are essentially the same as in this, though their details are complicated by the additional beam modulation necessary for making three-dimensional position estimates.

2.2.1 Position-dependent fluorescence rate modulation

As illustrated in Fig. 2.2, we let $\mathbf{x}_s(t)$ be the position of the tracking stage and $\mathbf{x}_p(t)$ be the position of the tracked particle at time t. For a Gaussian laser beam, the fluorescence intensity depends only on the relative coordinate $\mathbf{e}(t) = \mathbf{x}_p(t) - \mathbf{x}_s(t)$. If the beam is rotating at frequency ω_{xy} with rotation radius r, and we define the focused waist $w(0) = w_0$, we calculate the fluorescence intensity from Eq. 2.1:

$$\Gamma(\mathbf{e}, t) = \chi I(\mathbf{e}, t) = \Gamma_0 \exp\left\{-\frac{2}{w_0^2}\left[(e_x(t) - r\cos\omega_{xy}t)^2 + (e_y(t) - r\sin\omega_{xy}t)^2\right]\right\}, \qquad (2.3)$$

where χ is the fluorescence scattering coefficient of the particle and $\Gamma_0 = 2P\chi/\pi w_0^2$ is the fluorescence intensity of a particle located at the origin. We can express this simply in polar coordinates as

$$\Gamma(\mathbf{e}, t) = \Gamma_0 \exp\left\{-\frac{2}{w_0^2}\left[\rho^2 + r^2 - 2r\rho\cos(\omega_{xy}t - \phi)\right]\right\}, \qquad (2.4)$$

where, as usual, $\rho \equiv \|\mathbf{e}\|$ and $\phi \equiv \tan^{-1}[e_y/e_x]$.

Equation 2.4 gives us the quantitative form of the intuitive fluorescence rate modulation that we described earlier and illustrated in Figure 2.2. The size of the oscillating term in the exponent is proportional to ρ, and ϕ is exactly the phase angle of the oscillation. Before we try to make quantitative sense out of it, however, we first discuss the very general problem of demodulation of AC signals. This topic deserves its own sub-section because it is very important to the functioning of our apparatus.

2.2.2 Lock-in detection: demodulation of AC signals

In the previous section we derived Eq. 2.4, which we know contains information describing the two-dimensional position $\mathbf{x}_p$ in the magnitude and phase of its ω_{xy} oscillation. The question now is how to extract that information efficiently. Here we solve the general problem, where our signal consists of a superposition of N sinusoidal terms at distinct frequencies and we must measure the magnitude and phase of one of them. This procedure is referred to as *phase-sensitive* or *lock-in* detection, and is exactly how laboratory lock-in amplifiers work.

Specifically, if $u(t)$ is a signal that can be written as a finite sum of sine waves,

$$u(t) = \sum_{n=0}^{N} a_n \cos(\omega_n t - \phi_n), \qquad (2.5)$$

then we can extract the amplitude a_j and phase ϕ_j of any frequency component ω_j by multiplying by a sine wave with frequency ω_j and averaging over bandwidth B:

$$\begin{aligned} v(B) &= A_V B \int_0^{1/B} \mathrm{d}t \, \cos\left(\omega_j t\right) u(t) \\ &= \frac{A_V B}{2} \int_0^{1/B} \mathrm{d}t \sum_{n=0}^{N} a_n \left\{ \cos\left[\left(\omega_n - \omega_j\right) t - \phi_n\right] + \cos\left[\left(\omega_n + \omega_j\right) t - \phi_n\right] \right\}, \end{aligned} \tag{2.6}$$

where A_V is a gain that translates from the signal's units (Volts) to the units of v.

The frequency-difference term for which $\omega_n = \omega_j$ has no t dependence and contributes a value of $A_V a_j \cos(\phi_j)/2$ to the integral. For other frequencies ω_n, if $|\omega_n - \omega_j|/B \gg 2\pi$ then the cosine terms oscillate very quickly and their contribution to the average scales as $B/|\omega_n - \omega_j|$. Any reasonable choice of B will ensure that this is always true for the terms containing $\omega_n + \omega_j$, so those terms average out of the calculation; furthermore, we can make the contribution due to the ω_n terms arbitrarily small simply by making B smaller, and this has no impact on our measurement of the fixed value ϕ_j. In practical applications such as our tracking experiment, the information encoded by ϕ_j fluctuates fairly rapidly and these fluctuations serve, in effect, to broaden the linewidth of the ω_j frequency component of the signal. If B is set too small relative to this broadened linewidth these fluctuations will also average out of the integral.

The procedure that we have just described is useful when only one of a_j and ϕ_j contains information that we need to recover, because it is impossible to extract both terms simultaneously from the average

$$v(B) \approx \frac{A_V a_j}{2} \cos(\phi_j). \tag{2.7}$$

This is called single-phase detection because it returns only the cosine quadrature of ϕ_j. For our tracking application, information is simultaneously encoded in both a_j and ϕ_j, so we must perform lock-in detection on both quadratures (dual-phase detection) using

$$v(B) = \frac{A_V B}{2} \int_0^{1/B} \mathrm{d}t \, e^{-i\omega_j t} u(t) \approx \frac{A_V a_j}{2} e^{i\phi_j}. \tag{2.8}$$

This average is sufficient to independently recover both a_j and ϕ_j.

The most important result of this section is the fact that we can extract information from a *single* frequency out of a superposition of an any number of frequencies, with contamination from other frequencies scaling as $B/|\omega_n - \omega_j|$. In our tracking experiment, the amount of noise present in the fluorescence signal at exactly the rotation frequency ω_{xy} (which is typically several tens of kHz) is generally very small. Contributions to our estimate due to low-frequency noise sources discussed in Section 2.1 are attenuated by a factor of B/ω_{xy}, which is very small as we have already discussed. This implies that we can make position estimates that are almost completely free of systematic noise (although laser intensity noise appears in the a_j term, as we will discuss in the next section). Furthermore, we may encode additional information — such as the z position of the particle for three-dimensional tracking — at other high frequencies, and there will be hardly any cross-talk between the demodulated signals if those frequencies are sufficiently separated from ω_{xy}.

2.2.3 The two-dimensional localization signal

Before actually calculating what the localization signal is, we should consider what we want it to be. The output of our lock-in amplifiers will consist of a separate analog electronic signal for each Cartesian axis. Those signals should equal zero when the particle is on the axis of the beam rotation, and should vary linearly with the particle's distance from the origin so that it is easy to translate the electronic signals to real position estimates. Of course, this linearity must break down at some point, because we cannot localize the particle when it is far from the rotation axis since it will be only very weakly illuminated by the laser beam.

The fluorescence signal collected by our detector consists of a stream of photons

$$\xi(t) = \sum_{j}^{N} \delta(t - t_j), \tag{2.9}$$

where the arrival times t_j are random variables drawn from a point process[8] with modulated rate given in Eq. 2.4 and δ is the Dirac delta function. Each photon produces an electronic pulse $h(t)$ on the output of the detector that is then used for localization measurements. The complete electronic signal $u(t)$ is the convolution of $\xi(t)$ and $h(t)$, but for simplicity we assume that $h(t)$ is itself a sharply-peaked function that integrates to V_0. This way, on time-scales relevant to our localization estimate (time-scales comparable to or longer than ω_{xy}^{-1}), $h(t) \approx V_0 \delta(t)$ and we have

$$u(t) \approx V_0\, \xi(t). \tag{2.10}$$

We use lock-in detection to extract the magnitude and phase of the ω_{xy} component of $u(t)$; due to the stochastic nature of the detected fluorescence signal, this output is a random process. We must characterize this random process by computing its expected value, which will tell us the average value of the position estimate for a particle at position **e**. For this we must use the statistics of rate-modulated Poisson processes[8, 37]: if $\xi(t)$ is the derivative of a Poisson counting process (as defined in Eq. 2.9) and the rate of this process is $\Gamma(t)$, then

$$\langle \xi(t) \rangle = \Gamma(t) \tag{2.11}$$

$$\langle \xi(s)\xi(t) \rangle = \Gamma(s)\Gamma(t) + \Gamma(t)\delta(t - s). \tag{2.12}$$

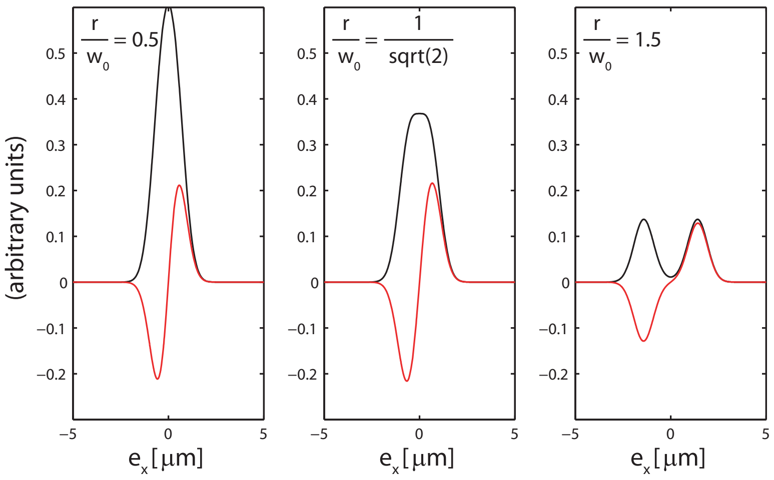


Figure 2.3: Beam intensity (black) and localization signals (red) for the specified rotation radii, with the beam waist w_0 fixed at 1μm.

Using Eq. 2.11 and the rate $\Gamma(\mathbf{e}, t)$ from Eq. 2.4, the expected value of the estimate $\hat{\mathbf{e}}$ is

$$\begin{aligned}\langle\hat{\mathbf{e}}\rangle &= A_V B V_0 \Gamma_0 \exp\left[-\frac{2}{w_0^2}(\rho^2+r^2)\right]\int_0^{1/B} \mathrm{d}t\, e^{-i\omega_{xy}t}\exp\left[\frac{4r\rho}{w_0^2}\cos(\omega_{xy}t-\phi)\right] \\ &\approx A_V V_0 \Gamma_0 \exp\left[-\frac{2}{w_0^2}(\rho^2+r^2)\right] I_1\left[\frac{4r\rho}{w_0^2}\right]e^{i\phi}, \end{aligned} \tag{2.13}$$

where I_1 is the modified Bessel function and the approximation is made by assuming that $B \ll \omega_{xy}/2\pi$. It is clear from the phase of Eq. 2.13 that the *direction* of $\hat{\mathbf{e}}$ is, on average, correct; to see that the magnitude of $\hat{\mathbf{e}}$ is proportional to ρ, we compute the Taylor expansion of Eq. 2.13 and get

$$|\langle\hat{\mathbf{e}}\rangle| = \frac{2A_V V_0 \Gamma_0 r}{w_0^2}\exp\left(-\frac{2r^2}{w_0^2}\right)\left[\rho + \frac{2r}{w_0^4}\left(\frac{r^2}{w_0^2}-1\right)\rho^3 + O(\rho^4)\right] \tag{2.14}$$

so that for $\rho < w_0$ (assuming $w_0 \sim r$), the linear term is dominant. This proportionality does not give us calibrated position estimates unless we carefully tune A_V so that

$$A_V = \frac{w_0^2}{2V_0\Gamma_0 r}\exp\left(\frac{2r^2}{w_0^2}\right), \tag{2.15}$$

but a well-designed feedback system should not depend on such calibration: our tracking system seeks simply to lock $\hat{\mathbf{e}} = 0$, without any regard for the size of the displacements needed to do so. However, calibration is necessary for relating the size of electronic noise to the size of position estimate fluctuations, as will be done in the next Section.

In Figure 2.3 we show a cross-section of the localization signal from Eq. 2.13 for a few different values of the rotation radius. If the radius is too big, the intensity profile has a donut shape and the slope of the localization signal is reduced at the origin. The approximately linear form of the signal near the origin as predicted in Eq. 2.14 is apparent in all three plots, as is the fall-off of the localization signal for large e_x due to the decaying Gaussian beam intensity.

We should note that the linearity criterion $\rho < w_0$ defines what is considered the "good" regime for localization; most of the analysis in this and the next chapter assumes that this inequality holds while tracking. However, if a particle moves too quickly for our localization optics to track it or for our tracking stage to follow it, we must consider the consequences of the nonlinear terms. We do so in Section 3.3 of the next chapter.

We should also note that laser intensity noise will cause fluctuations in Γ_0, creating a multiplicative noise term in Eq. 2.14. Provided that certain assumptions are true — that the noise is a small fraction of Γ_0 and is not at the rotation frequency of the laser beam — the noise can never displace the location of the *zero* in the localization signal or cause it to point in the wrong direction. Instead, these fluctuations feed into the loop bandwidth of the tracking system, causing fluctuations in the tracking error statistics. However, such fluctuations will average out on time-scales much shorter than the tracking system bandwidth, and can be suppressed using feedback to control the intensity on time-scales typically extending well beyond the feedback bandwidth.

2.2.4 Two-dimensional localization noise

We can only collect relatively small numbers of photons from single fluorescent nanoparticles in any finite time $1/B$, and as a result there will always be photon-counting shot noise on our position estimates. In the previous section we showed that our estimator $\hat{\mathbf{e}}$ will always approximate the position $\mathbf{e}$ of a stationary particle if we average for a long time. However, to track a *moving* particle we would like to be able to average for as short a time as possible. In general, a compromise must be reached in setting the tracking bandwidth, because averaging for too long will prevent us from tracking fast-moving particles, but not averaging for long enough will inject excess shot noise into the position estimates. We will now compute the position estimation noise on $\langle\hat{\mathbf{e}}\rangle$ resulting from photon-counting statistics.

Using Eqs. 2.4 and 2.12 we can calculate the variance

$$\begin{aligned}\left\langle |\hat{\mathbf{e}} - \langle\hat{\mathbf{e}}\rangle|^2 \right\rangle &= A_V^2 B^2 \int_0^{1/B} \mathrm{d}s \int_0^{1/B} \mathrm{d}t\, e^{i\omega_{xy}(t-s)} \langle u(s)u(t)\rangle - \langle\hat{\mathbf{e}}\rangle^\dagger \langle\hat{\mathbf{e}}\rangle \\ &\approx A_V^2 B V_0^2 \Gamma_0 \exp\left[-\frac{2}{w_0^2}\left(\rho^2 + r^2\right)\right] I_0\left[\frac{4r\rho}{w_0^2}\right] \qquad (2.16)\end{aligned}$$

which tells us, as a function of B, the degree of uncertainty in our position estimate due simply to the fact that we collect a finite number of photons in the averaging interval B^{-1}. Since Eq.

2.16 is independent of ϕ, it represents the localization noise density along each Cartesian axis.

In the regime $\rho < r$ where the tracking error signal is a linear function of ρ we have

$$\left\langle |\hat{\mathbf{e}} - \langle \hat{\mathbf{e}} \rangle|^2 \right\rangle \approx A_V^2 B V_0^2 \Gamma_0 \exp\left[-\frac{2r^2}{w_0^2}\right] \tag{2.17}$$

$$= \frac{w_0^4}{4r^2}\frac{B}{\Gamma_0}\exp\left[\frac{2r^2}{w^2}\right], \tag{2.18}$$

where Eq. 2.18 gives us the variance in terms of the position estimate rather than an electronic signal, and comes from inserting Eq. 2.15 into Eq. 2.17. We may go one step further and compute the signal-to-noise ratio

$$\frac{|\langle \hat{\mathbf{e}} \rangle|}{\sqrt{\left\langle |\hat{\mathbf{e}} - \langle \hat{\mathbf{e}} \rangle|^2 \right\rangle}} = \frac{2r}{w_0^2}\sqrt{\frac{\Gamma_0}{B}}\exp\left(-\frac{r^2}{w_0^2}\right)\rho. \tag{2.19}$$

This expression indicates that the signal-to-noise ratio depends only on the beam and rotation geometry, the fluorescence rate and integration bandwidth, and the position of the particle from the center of the rotation. At a fixed fluorescence rate and integration bandwidth, we can maximize the signal-to-noise ratio by setting $r = w_0/\sqrt{2}$. As shown in Fig. 2.3, this is the largest radius possible before a donut-shaped hole appears in the center of the excitation profile. This gives us a guideline to follow in constructing the apparatus: for a given beam waist, we would like our rotation radius to be close to $w_0/\sqrt{2}$ in order to make the best possible localization estimates. It may seem tempting to try to intuit the relationship between r and w_0 in the optimal configuration. For example, we might guess that r and w_0 might maximize the slope of the localization signal in Eq. 2.14, or position the beam so that the most sensitive spot in its Gaussian profile is always at the center of the rotation. The former of these occurs at $r = w_0\sqrt{3}/2$, however, and the latter at $r = w_0/2$. This is not to say that there is no intuitive explanation for the optimal geometry, I just do not know what it is.

The noise figure given in Eq. 2.18 represents an absolute limit to how well we may localize any fixed fluorescent particle while averaging for time $1/B$. It contributes an estimation error that feeds into our tracking stage positions and reduces tracking accuracy. In order to reduce this noise, we would ideally choose smaller averaging bandwidths; however, doing this while tracking a *moving* particle allows the particle to travel further, adding more uncertainty to the position estimate. We do not discuss this here, but Andy Berglund's work [36, 37] contains a thorough discussion on this topic.

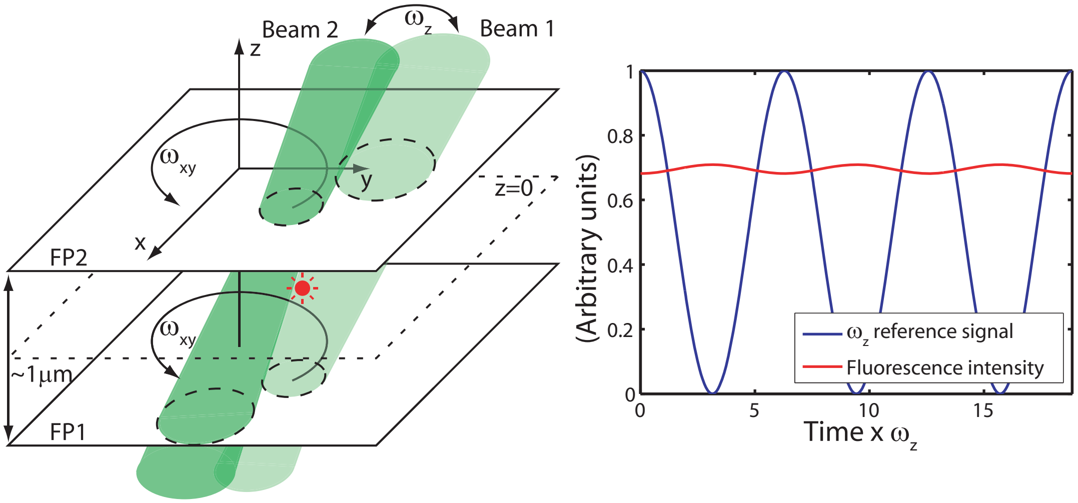


Figure 2.4: **Left**: Modulated beam geometry for three-dimensional localization. Two rotating laser beams are focused in two different focal planes (FP1 and FP2), and the optical power in the two beams is modulated 180° out of phase at frequency ω_z. **Right**: Exaggerated fluorescence intensity modulation due to particle's offset along the z axis. Under typical experimental conditions, even for large displacements on the z axis, this size of this modulation is usually about 1% what is shown, or 0.1% of the DC fluorescence intensity.

2.3 Three-dimensional localization

In the previous section, we overcame the symmetry of the excitation beam by rotating it, inducing temporal asymmetry that we then used to extract two-dimensional position estimates from the fluorescence signal. By itself, this procedure does nothing to help us estimate the position of the particle along the z axis, but essentially the same idea can be used to solve this problem. By moving the rotating beam up and down at a frequency $\omega_z \neq \omega_{xy}$, we can encode the z position of the particle in a separate frequency component of the fluorescence signal and recover it using a second lock-in detector.

Enrico Gratton's group [11, 32, 35] was the first to implement an approach like this. They used a piezoelectric nanopositioner stage to move their microscope objective up and down at a frequency of about 100Hz, and they extracted position estimates using an algorithm based on FFT. Using this method they demonstrated tracking of fluorescent spheres inside of living cells, with an axial tracking bandwidth of 3Hz. The tracking bandwidth in their case was limited by the relatively low modulation frequency: as we showed in Section 2.2.2, the averaging bandwidth B must be smaller than the modulation frequency, but B also sets the maximum tracking bandwidth. This relatively slow bandwidth is not sufficient for tracking fast-moving objects such as single molecules, but it is difficult to precisely move a heavy microscope objective much faster than 100Hz; an alternative approach is needed.

Figure 2.4 illustrates the axial localization method that we have developed. Two laser beams rotating at frequency ω_{xy} are focused at different depths inside the sample, separated by about 1μm, and the total excitation power is alternated between the beams at frequency ω_z. As in two-dimensional localization, the particle's position in the xy plane is encoded in the magnitude and phase of the ω_{xy} frequency component of the fluorescence signal. The particle's position in z is encoded in the signal in a similar way: as shown in the plot in Fig. 2.4, the ω_z frequency component of the fluorescence signal is either in-phase or 180° out-of-phase with the ω_z drive signal, depending on whether the particle is above or below $z = 0$. The magnitude of the ω_z component is proportional to the particle's distance to the origin. Optical power modulation can easily be done at high frequencies: we typically use $\omega_z = 2\pi \times 100$kHz, with a demodulation bandwidth $B \sim 1$kHz. This bandwidth will almost never be the limiting factor in single-particle tracking applications, because photon-counting noise will typically place a much lower limit on the localization bandwidth.

Using a pair of beams in three dimensions introduces several free parameters describing the beam geometry that are not present in two dimensions: each beam is focused in a different plane, with a different minimum waist. In addition to the waist of the beam, the rotation radius r also depends on z, forming a cone that focuses down so that $r(z_0) = 0$ at some depth z_0. Before characterizing the localization statistics of this method quantitatively, we will describe the beam geometry in further detail.

2.3.1 More on beam geometry

To generally describe the two localization beams we must keep track of the z-dependent waist and rotation radius of both beams separately, but this would produce extremely complicated mathematical descriptions that, in practical terms, are unnecessary. Instead, we will assume that the focused waists w_0 of the two beams are identical and that the angle of divergence of the cone of rotation is identical for the two beams. They are then allowed to differ only in whether the beam waist focuses to its minimum above or below the rotation cone. In practice, it should be possible to create a pair of beams that are very close to identical, so this assumption is not overly restrictive.

Figure 2.5 illustrates two beam geometries that fit within these specifications. In the first geometry, the rotation cone of beam 1 focuses below its beam waist, and the rotation cone of beam 2 focuses above its beam waist. Because of its symmetry about $z = 0$, this geometry produces a localization signal that is symmetric and goes to zero at exactly $z = 0$. In the second geometry, the waists of both beams focus on the same side of the rotation cone foci. The asymmetry of this geometry about $z = 0$ translates to asymmetry in the localization signal and causes the zero of the localization estimate to occur in a plane $z \neq 0$. In both geometries,

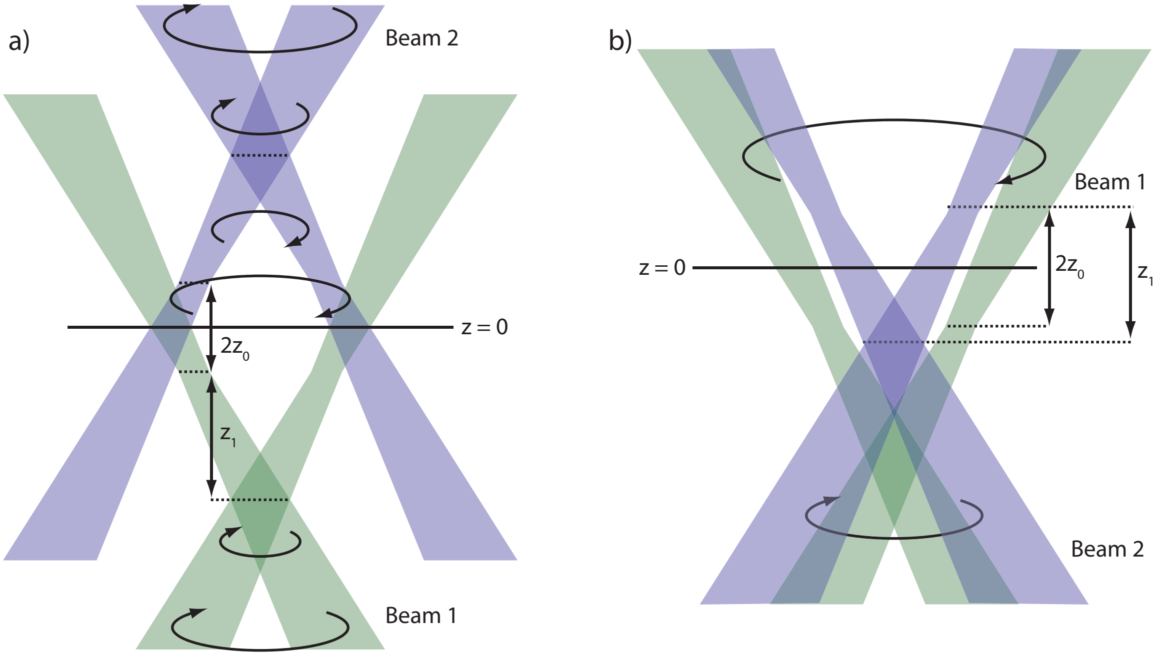


Figure 2.5: Cross-sections of the three-dimensional excitation profiles for two modulated beam geometries. **a**: geometry symmetric about the tracking plane. **b**: an asymmetric geometry. Different colors are used to clearly distinguish between the two beams, which are the same color in the experiment. Each beam appears twice, on either side of the rotation orbit. In both figures, $z = 0$ indicates the plane that is half-way between the focal planes of the waists of the two beams. Both figures are drawn to an exaggerated scale for clarity; in the actual apparatus, the divergence of the rotation cone is much smaller. Note that the beam rotation is shifted by 180° on opposite sides of the plane $r(z) = 0$, as suggested by the differently-oriented arrows.

the distance between the planes in which the beam waists are focused is denoted $2z_0$, and the distance between the foci of the beam waist and rotation radii for either beam is denoted z_1.

In either geometry the time-dependent fluorescence intensity distribution is given by

$$\Gamma(\mathbf{e},t) = \frac{w_0^2\Gamma_0}{2w(e_z+z_0)^2}(1-\cos\omega_z t)\exp\left[-2\frac{\rho^2+r_1(e_z)^2-2\rho r_1(e_z)\cos(\omega_{xy}t-\phi)}{w(e_z+z_0)^2}\right]$$
$$+\frac{w_0^2\Gamma_0}{2w(e_z-z_0)^2}(1+\cos\omega_z t)\exp\left[-2\frac{\rho^2+r_2(e_z)^2-2\rho r_2(e_z)\cos(\omega_{xy}t-\phi)}{w(e_z-z_0)^2}\right], \quad (2.20)$$

where the beam waists $w(z)$ are given by Eq. 2.2. The radii of rotation $r_1(z)$ and $r_2(z)$ are linear functions of z and can, in principle, be varied independently of z_0 by appropriate alignment of the experimental apparatus. The exact expressions for the radii depend on the beam geometry and will be provided for the geometries we have discussed in the next sections.

To extract an axial localization signal $\hat{e}_z$ from this general expression for the fluorescence intensity we use single-quadrature lock-in detection, which we evaluate just as we did in Section 2.2.3 to get

$$\langle\hat{e}_z\rangle = \frac{A_V V_0 \Gamma_0 w_0^2}{4}\left\{\frac{1}{w(e_z-z_0)^2}\exp\left[-2\frac{\rho^2+r_2(e_z)^2}{w(e_z-z_0)^2}\right]I_0\left[\frac{4\rho r_2(e_z)^2}{w(e_z-z_0)^2}\right]\right.$$
$$\left.-\frac{1}{w(e_z+z_0)^2}\exp\left[-2\frac{\rho^2+r_1(e_z)^2}{w(e_z+z_0)^2}\right]I_0\left[\frac{4\rho r_1(e_z)}{w(e_z+z_0)^2}\right]\right\}. \quad (2.21)$$

If we now take the Taylor expansion about $\rho = 0$, since we assume that the tracking system keeps the particle close to the origin in those coordinates, we get

$$\langle\hat{e}_z\rangle = \frac{A_V V_0 \Gamma_0 w_0^2}{4}\left\{\frac{1}{w(e_z-z_0)^2}\exp\left[-\frac{2r_2(e_z)^2}{w(e_z-z_0)^2}\right]\right.$$
$$\left.-\frac{1}{w(e_z+z_0)^2}\exp\left[-\frac{2r_1(e_z)^2}{w(e_z+z_0)^2}\right]\right\}+O(\rho^2), \quad (2.22)$$

which is an important result: small displacements along the x and y axes do not couple into the position estimates for the z axis. This general form of the z position estimate can be further simplified if we now parameterize the beam geometry more specifically.

2.3.2 Localization signal for the symmetric geometry

In the case of the symmetric focal volume geometry shown in Fig. 2.5, the rotation radii of the two beams are given by $r_1(e_z) = r_0|e_z + z_0 + z_1|/z_1$ and $r_2(e_z) = r_0|e_z - z_0 - z_1|/z_1$. In both cases, we have defined r_0 as the rotation radius of the beam in the planes $z = \pm z_0$. Using these

	w_0 [μm]	r_0 [μm]	z_0 [μm]	z_1 [μm]
a	0.5	0.35	2	5
b	0.5	0.35	2	2
c	0.5	0.5	5	5
d	0.5	0.35	2	∞
e	2	1.414	2	5
f	0.5	0.7	2	5

Table 2.1: Parameters used in the plots in Fig. 2.6

expressions, we approximate Eq. 2.22 to first order in e_z to get

$$\langle \hat{e}_z \rangle \approx \frac{A_V V_0 \Gamma_0 w_0^4}{w(z_0)^4 z_R^2} \exp\left[-\frac{2r_1(0)^2}{w(z_0)^2}\right] \left\{ \underline{z_0 + \frac{2r_0 r_1(0)}{w(z_0)^2 z_1}\left[z_R^2 - z_0 z_1\right]} \right\} e_z, \tag{2.23}$$

which has the linearity in e_z that makes for a good axial localization signal. However, it is not generally true that $\langle \hat{e}_z \rangle / e_z > 0$ for all z_0, z_1, z_R, w_0 and r_0 because the underlined term in Eq. 2.23 is not bounded from below. This means that for some combinations of these parameters, $\hat{e}_z$ will point in the wrong direction. It is easy to see how this can happen if we take imagine a rotation radius that does not depend on z (equivalent to taking $z_1 \to \infty$). In such a case, if $w_0 < r_0\sqrt{2}$, then a particle located on the optic axis experiences a donut-shaped excitation intensity near z_0. The central hole in the donut gets smaller with distance from either beam's focal plane because of expansion of the beam waist. Thus, for a particle near, but not at, $e_z = 0$, the beam that illuminates it most brightly may be that which has the smaller central hole, not the smaller waist. Fortunately, this problem is limited to a pathological case: we can show using SOSTOOLS[38] that for $r_0 \leq w_0/\sqrt{2}$, the underlined term is a sum-of-squares and hence nonnegative for all nonnegative z_0, z_1 and z_R.

2.3.3 Localization signal for the asymmetric geometry

In the asymmetric geometry of Figure 2.5, the rotation radii are given by $r_1(e_z) = r_0|e_z + z_1 + z_0|/z_1$ and $r_2(e_z) = r_0|e_z - z_0 + z_1|/z_1$. As we discussed previously, the localization signal is generally nonzero at $z = 0$, rather going to zero in some other plane that is the solution to Eq. 2.22 with these rotation radii. It is not easy to derive an expression for the location of this plane, and therefore we cannot easily compute a Taylor expansion of $\langle \hat{e}_z \rangle$ to show that it is proportional to e_z. Instead, we compute the localization signal numerically for several values of the parameters w_0, r_0, z_0 and z_1, and we show the results in Fig. 2.6. The values of the parameters used in the figure are given in Table 2.1.

Figure 2.6 shows that the localization signals for the asymmetric geometry appear very similar to those for the symmetric geometry for most choices of parameters. We show the localization signals for a range of parameter values, in order to illustrate the effects of the

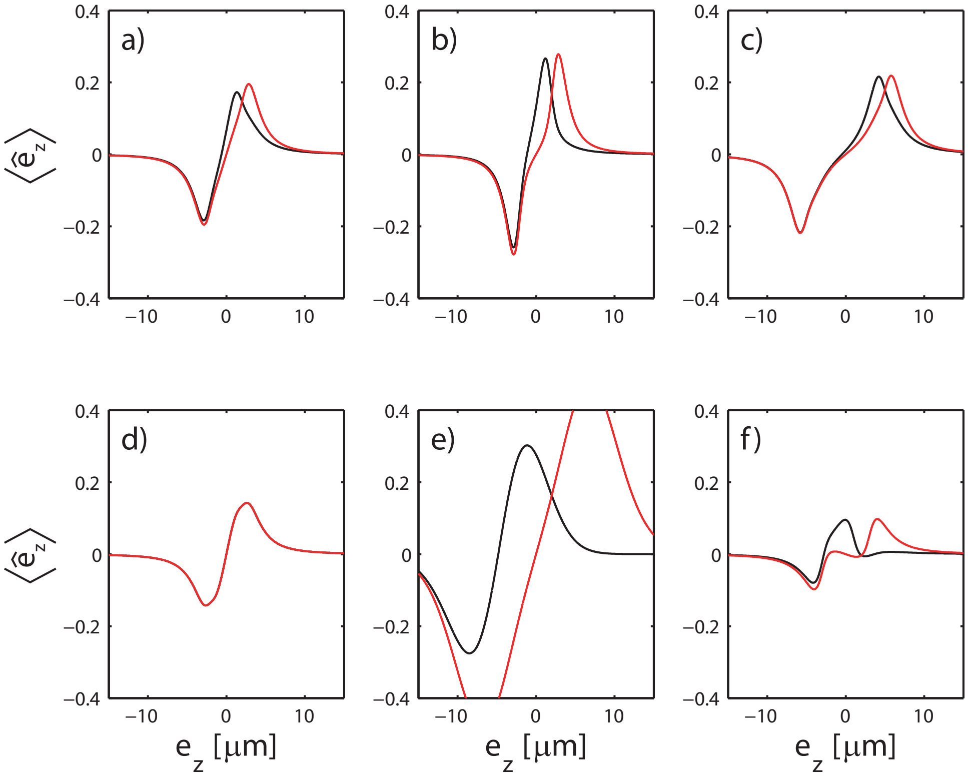


Figure 2.6: Localization signals for the axial position estimate from Eq. 2.22 for the symmetric (red) and asymmetric (black) beam geometries. We used $A_V V_0 \Gamma_0 = 1$ in all plots; the remaining parameters are given in Table 2.1.

parameters — including when their values are chosen poorly. Plots **a**-**c** show the effect of varying z_0 and z_1: when z_1 is made small, resulting in wider divergence of the rotation cone, or z_0 is made large, the localization signal loses sensitivity at the origin. Plot **d** illustrates the equivalence of the two geometries when $z_1 \to \infty$, meaning the rotation radius does not depend on z. Plot **e** shows that increasing w_0 (and, in proportion, r_0) yields a greater difference between geometries, and for both geometries it yields error signals over a wider range in z (the difference in the size of the error signals is an artifact of the factor of w_0^{-2} hidden inside Γ_0). Plot **f** is an example of a pathological configuration, in which the rotation radius is too large. Here both signals are small and misshapen, and the symmetric geometry yields a signal that points in the wrong direction at the origin.

2.3.4 Radial localization with three-dimensional modulation

We have not yet considered what happens to the two-dimensional localization signal as a consequence of our three-dimensional modulation. The simple cosine dependence of Eq. 2.20 on ω_z makes it clear that dual-phase lock-in detection on the ω_{xy} component returns (as in Eq. 2.13)

$$\begin{aligned}\left\langle \hat{\mathbf{e}}_{xy} \right\rangle = \frac{A_V V_0 \Gamma_0 w_0^2}{2} \Bigg\{ & \frac{1}{w(e_z+z_0)^2} \exp\left[-\frac{2\left(\rho^2 + r_1(e_z)^2\right)}{w(e_z+z_0)^2} \right] I_1\left[\frac{4 r_1(e_z)\rho}{w(e_z+z_0)^2} \right] \\ & + \frac{1}{w(e_z-z_0)^2} \exp\left[-\frac{2\left(\rho^2 + r_2(e_z)^2\right)}{w(e_z-z_0)^2} \right] I_1\left[\frac{4 r_2(e_z)\rho}{w(e_z-z_0)^2} \right] \Bigg\} e^{i\phi}. \end{aligned} \tag{2.24}$$

We know that the derivative of $\left\langle \hat{\mathbf{e}}_{xy} \right\rangle$ must equal zero in the symmetric beam geometry. Inserting the expressions for the beam waist and rotation radii we get

$$\left\langle \hat{\mathbf{e}}_{xy} \right\rangle = \frac{A_V V_0 \Gamma_0}{w(z_0)^2} \exp\left[-\frac{2}{w(z_0)^2} \left(r_1(0)^2 + \rho^2 \right) \right] I_1\left[4 \frac{r_1(0)\rho}{w(z_0)^2} \right] e^{i\phi} + O(e_z^2). \tag{2.25}$$

This result is identical to the localization signal with two-dimensional modulation given in Eq. 2.13; our additional modulation along the z-axis has no effect on the localization signal.

For the asymmetric geometry it will generally *not* be the case that the estimate of $\hat{\mathbf{e}}_{xy}$ is identical to Eq. 2.13. Explicit calculation of this dependence is horribly messy and not particularly informative, so we will not offer it here. However, due to the similarity between the localization signals for the symmetric and asymmetric geometries as shown in Fig. 2.6, we suspect that this e_z dependence will be very small for most combinations of beam parameters.

2.3.5 Three-dimensional localization noise

In Section 2.2.4 we discussed Andy Berglund's calculations of the statistical uncertainty in two-dimensional localization estimates arising from the counting noise present when working with small numbers of photons. Here we evaluate the size of this noise on our localization estimates in three-dimensions. The approach taken in this section is identical to that in Section 2.2.4, so we move quickly through the calculations. More detail can be found in that section.

Considering our results in the previous section, it is not surprising that the noise on the radial localization estimate $\left\langle \hat{\mathbf{e}}_{xy} \right\rangle$ with symmetric beam geometry is the same as that with two-dimensional modulation (Eq. 2.17):

$$\left\langle \left| \hat{\mathbf{e}}_{xy} - \left\langle \hat{\mathbf{e}}_{xy} \right\rangle \right|^2 \right\rangle = \frac{A_V^2 B V_0^2 \Gamma_0 w_0^2}{w(z_0)^2} \exp\left[-\frac{2 r_1(0)^2}{w(z_0)^2} \right] + O(\rho^2) + O(e_z^2). \tag{2.26}$$

This tells us that the radial localization noise is optimized by the same combination of beam geometry parameters as in two dimensions. Specifically, we should aim to set $r(0) = w(z_0)/\sqrt{2}$ in aligning our apparatus.

The noise on the estimate for e_z, for either of the beam geometries we have discussed, can be written as

$$\begin{aligned}\left\langle (\hat{e}_z - \langle \hat{e}_z \rangle)^2 \right\rangle = \frac{A_V^2 B \Gamma_0 V_0^2 w_0^2}{4} \Bigg\{ & \frac{1}{w(e_z + z_0)^2} \exp\left[-\frac{2 r_1(e_z)^2}{w(e_z + z_0)^2} \right] \\ & + \frac{1}{w(e_z - z_0)^2} \exp\left[-\frac{2 r_2(e_z)^2}{w(e_z - z_0)^2} \right] \Bigg\} + O(\rho^2). \end{aligned} \tag{2.27}$$

We can expand this expression to first order in e_z if we insert the parameters for the symmetric geometry. This gives us

$$\left\langle (\hat{e}_z - \langle \hat{e}_z \rangle)^2 \right\rangle = \frac{A_V^2 B \Gamma_0 V_0^2 w_0^2}{2 w(z_0)^2} \exp\left[-\frac{2 r_1(0)^2}{w(z_0)^2} \right] + O(\rho^2) + O(e_z^2). \tag{2.28}$$

Ignoring the second-order dependence on ρ and e_z, we finally compute the signal-to-noise ratio for the z-axis localization, accurate to first order in e_z and ρ:

$$\frac{\langle \hat{e}_z \rangle}{\sqrt{\left\langle (\hat{e}_z - \langle \hat{e}_z \rangle)^2 \right\rangle}} \approx \frac{w_0^3}{w(z_0)^3 z_R^2} \sqrt{\frac{2\Gamma_0}{B}} \exp\left[-\frac{r_1(0)^2}{w(z_0)^2} \right] \left\{ z_0 + \frac{2 r_0 r_1(0)}{w(z_0)^2 z_1} \left[z_R^2 - z_0 z_1 \right] \right\} e_z. \tag{2.29}$$

The general problem of finding the optimal geometry to maximize the signal-to-noise ratio for z-axis localization is difficult because of the complicated form of Eq. 2.29. In principle, we can fix certain parameters in order to first optimize the signal-to-noise ratio along the x and y axes, leaving fewer free parameters and a simpler optimization problem; however, precise control of the beam geometry is difficult, so such calculations would have little practical value.

Chapter 3

Tracking system dynamics

In the last chapter we discussed the optical method that we use to compute estimates of a fluorescent particle's position. This discussion focused entirely on static properties of the localization system; we did not consider the fact that the particle is moving and the tracking stage is following it. This chapter takes a very general result from the previous chapter — the fact that we have a well-characterized, accurate method to estimate the position of a fluorescent particle — and builds a feedback loop around it. We account for the dynamics of the particle's motion and the statistical properties of the localization estimate, showing how they affect the statistics of the tracking stage. In particular, we concern ourselves with the statistics of tracking errors — the deviations between the stage and particle positions. When these errors are small relative to the laser beam rotation radius, we are able to characterize them analytically in almost exact detail. When the errors are not so small — typically due to a particle that moves too fast for the feedback system to keep up — the localization estimates lose fidelity and the resultant tracking statistics are much more complicated. The chapter concludes with a discussion of some of the consequences of these larger errors and strategies that may be used to avoid them.

3.1 The feedback loop

Figure 3.1 shows a block diagram representation of the closed-loop feedback system. The particle's position $\mathbf{x}_p$ is the primary input. The position of the tracking stages is denoted by the variable $\mathbf{x}_s$, which corresponds precisely to the origin of the localization system described in Chapter 2. The displacement of the particle from the tracking stage (the error, $\mathbf{e} = \mathbf{x}_p - \mathbf{x}_s$) is the relevant input to the localization estimation method. We group all of the localization machinery — the laser beams, detectors, and lock-in amplifiers — conveniently into a single block that maps $\mathbf{e}$, which we cannot directly measure, into the estimate $\hat{\mathbf{e}}$ based on the measured fluorescence signal. As discussed in Section 2.2.4, $\hat{\mathbf{e}}$ is a noisy estimate, the inaccuracy of which is due to photon counting statistics. Rather than grouping this noise into the localization block,

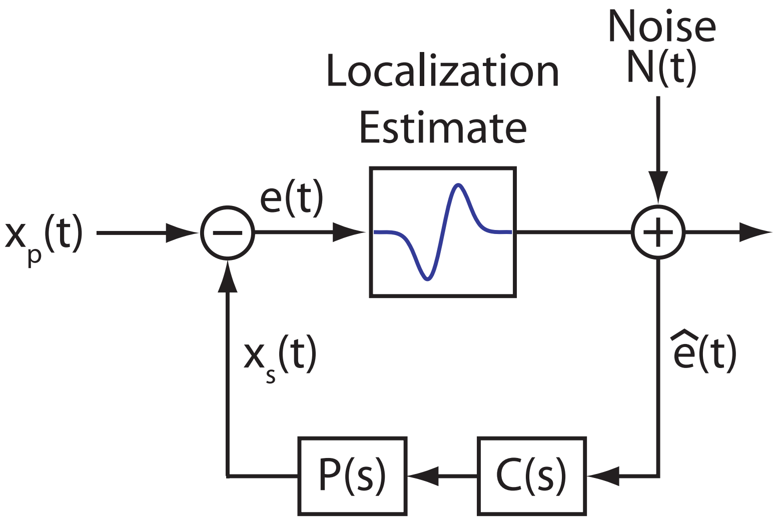


Figure 3.1: Tracking system feedback loop. The tracking error $\mathbf{e}(t)$ — the difference between the particle's position $\mathbf{x}_p(t)$ and stage position $\mathbf{x}_s(t)$ — causes fluorescence rate modulation that is used to estimate $\mathbf{e}(t)$. The error estimate $\hat{\mathbf{e}}(t)$ is fed into the feedback controller (with transfer function $C(s)$), and this drives the tracking stages (which have transfer function $P(s)$) in order to cancel the tracking error.

we feature it as an additive input $\mathbf{N}(t)$ to $\hat{\mathbf{e}}$ so that we can compute its effects on the tracking system dynamics later in this chapter.

Once $\hat{\mathbf{e}}$ has been constructed it is fed into a feedback controller, denoted by the $C(s)$ block in the diagram. The output of this controller is fed into the "plant" $P(s)$, consisting of the combination of the tracking stages and the electronic amplifiers that drive them. The stage positions feed back subtractively into $\mathbf{e}$, hence closing the loop.

The feedback loop is designed to drive $\hat{\mathbf{e}}$ to zero; since $\hat{\mathbf{e}}$ is an unbiased estimate of $\mathbf{e}$, this implies that $\mathbf{e}$ is, on average, zero as well. However, since $\hat{\mathbf{e}}$ is related to $\mathbf{e}$ by the addition of noise, enforcing $\hat{\mathbf{e}}(t) = 0$ corresponds to $\mathbf{e}(t) = \mathbf{N}(t)$, so that the localization noise feeds directly into tracking errors. Of course, all tracking systems have finite bandwidth so it is not possible to achieve $\hat{\mathbf{e}}(t) = 0$; as a result, both the localization noise and the motion of the particle contribute to the tracking error.

Figure 3.2 illustrates a simulated trajectory of a single particle and the corresponding trajectories of the tracking stage with varied feedback bandwidths. We hope this figure motivates the need to understand the statistics of the tracking errors. The trajectory taken by our tracking stage is a coarse approximation to the particle's actual trajectory; exactly how coarse depends on a variety of factors. An understanding of the fluctuations of x_p about x_s is necessary for describing fluorescence statistics in the next chapter and for determining the performance limits (in terms of how fast-moving a particle may be tracked) of the apparatus. It is our goal in the remainder of this chapter to provide a complete description of these statistics.

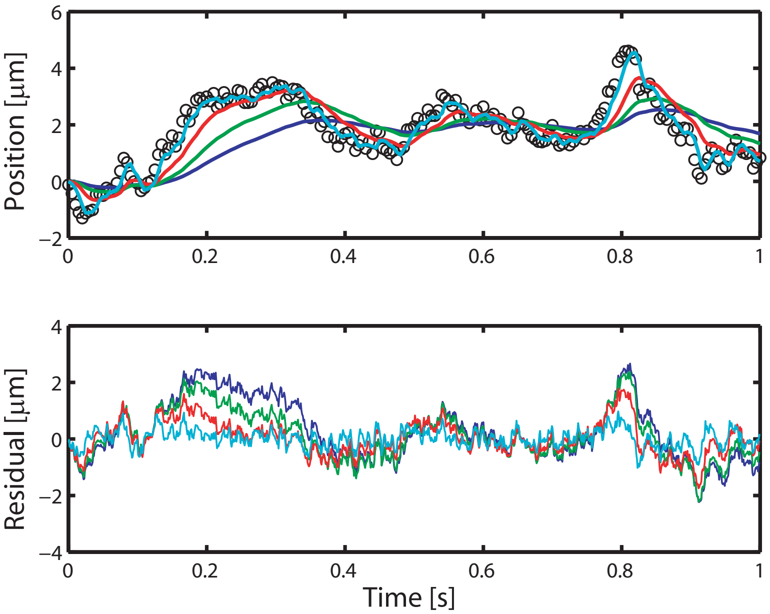


Figure 3.2: **Top**: Simulation of a diffusing particle (black circles) with diffusion coefficient $D = 10\mu\mathrm{m}^2/s$. We compute the trajectories of the tracking stage that are exact up to a random value for the initial state of the stage. The parameters describing the tracking system are defined in later sections; we used a second-order system with $\gamma_p = 100\mathrm{Hz}$ and controller bandwidth $\gamma_c = 1\mathrm{Hz}$ (blue), 2Hz (green), 5Hz (red) and 20Hz (cyan). **Bottom**: tracking error $e = x_p - x_s$.

3.1.1 Brownian motion

Before we go on to characterize the tracking system, we must have an adequate description of its inputs and constituent blocks. We begin with $\mathbf{x}_p$, following the Langevin derivation of the Brownian motion. For more information about this approach, see [8, 39, 40]. An excellent historical discussion of Brownian motion can be found in [40].

As discussed in Chapter 1, the particle's motion is caused by collisions between it and the molecules in the solvent surrounding it. All of the solvent molecules are in constant motion due to their thermal energy. Each time one of these molecules collides with the particle, the particle experiences a very small impulsive force pushing it in the direction the molecule was moving in. The thermal motion is uncorrelated between different solvent molecules, so that each collision pushes the particle in a different direction. We assume that the liquid molecules are very small relative to the particle, and that the mean free path between the molecules and the particle is very small. Therefore, the particle experiences a very large number of collisions in any short period of time.

There is no natural coordinate system for $\mathbf{x}_p$ because of the symmetry of the collisions with the solvent: all directions of those collisions are equally likely due to the equipartition of energy among the three translational degrees of freedom of the solvent molecules. Once a coordinate system is chosen, the particle's motion all three axes may be treated independently; there is no correlation between motion along orthogonal axes. As a result, henceforth we deal only with the scalar x_p representing any of the three components of $\mathbf{x}_p$.

The force exerted on the particle along x_p is a rapidly-fluctuating function that we denote $B(t)$. We write the particle's equation of motion as

$$m_p \frac{\mathrm{d}^2}{\mathrm{d}t^2} x_p = -\gamma \frac{\mathrm{d}}{\mathrm{d}t} x_p + B(t), \tag{3.1}$$

where m_p is the mass of the particle and the γ term represents the Stokes drag on the particle due to the viscosity of the solution. We must specify $B(t)$ in order to make any sense out of Eq. 3.1. The only tractable description of $B(t)$ is as a stochastic process — anything else would require that we keep track of the dynamics of all of the molecules in the solvent — so we characterize $B(t)$ by its statistical properties. First, $\langle B(t) \rangle = 0$ because we assume that there is no convective drift in the particle's position. Second, we let $B(t)$ be delta-correlated because each collision is very short:

$$\langle B(t_1) B(t_2) \rangle = \Upsilon \delta(t_2 - t_1), \tag{3.2}$$

where Υ is a constant that we will determine by physical arguments once we solve Eq. 3.1. This correlation time of zero characterizes $B(t)$ as *white noise*, because it implies (via the Wiener-Khinchin theorem) that $B(t)$ has a constant power spectral density. Finally, we assume that the distribution of $B(t)$ is Gaussian with zero mean and variance Υ. Figure 3.3**a** illustrates a realization of the process $B(t)$ simulated according to this description, with physical parameters chosen to correspond to a polystyrene sphere in water.

We may integrate Eq. 3.1 easily because it is linear, and we do so to get

$$v_p(t) \equiv \frac{\mathrm{d}}{\mathrm{d}t} x_p = e^{-\gamma t/m_p} v_p(0) + \frac{1}{m_p} e^{-\gamma t/m_p} \int_0^t \mathrm{d}\tau \, e^{\gamma \tau/m_p} B(\tau), \tag{3.3}$$

where we defined the velocity v_p for notational convenience. Figure 3.3**b** shows the particle's velocity computed from the simulated $B(t)$. Using Eq. 3.3, we may compute the mean and correlation function of the velocity:

$$\langle v_p(t) \rangle = e^{-\gamma t/m_p} v_p(0) \tag{3.4}$$

$$\langle\langle v_p(t_1) v_p(t_2) \rangle\rangle = \frac{\Upsilon}{2\gamma m_p} \left[e^{-\gamma |t_1 - t_2|/m_p} - e^{-\gamma (t_1 + t_2)/m_p} \right], \tag{3.5}$$

where we have used van Kampen's notation for the covariance[8], $\langle\langle v_p(t_1) v_p(t_2) \rangle\rangle \equiv \langle v_p(t_1) v_p(t_2) \rangle - \langle v_p(t_1) \rangle \langle v_p(t_2) \rangle$. Statistical thermodynamics gives us physical grounds for determining the steady-state variance of the velocity. It tells us that the probability distribution for the steady-

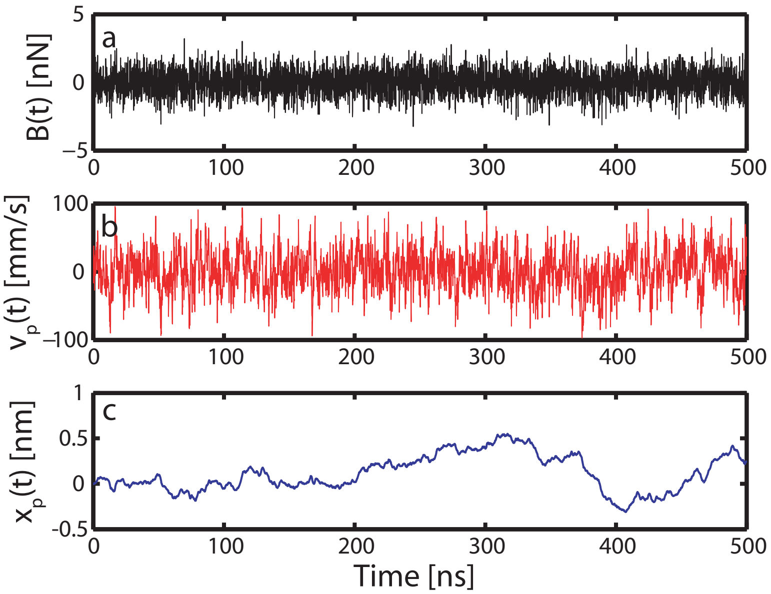


Figure 3.3: Simulated Brownian motion of a 1μm polystyrene sphere in water. We simulated the random force $B(t)$ (**a**, top) and computed the resulting particle velocity $v_p(t)$ (**b**, middle) and position $x_p(t)$ (**c**, bottom). Simulation parameters: $r = 0.5\mu$m, $\eta = 10^{-3}$ Pa·s (water), $\rho_p = 1.005$g/cm^3 (polystyrene), $T = 298$K.

state energy of the particle is exponential, with mean $k_B T$ where k_B is Boltzmann's constant and T is the temperature of the liquid. It is simple to show that this implies that v_p, related to the energy by $E = m_p v_p^2/2$, obeys

$$\lim_{t\to\infty} \langle v_p(t)^2 \rangle = \frac{k_B T}{m_p}. \tag{3.6}$$

We use this to determine the size of the fluctuations in $B(t)$ from Eq. 3.5:

$$\Upsilon = 2k_B T\gamma. \tag{3.7}$$

Due to our assumption that $B(t)$ is Gaussian, the higher cumulants of v_p all vanish. In other words, v_p is itself Gaussian, so that it is fully characterized by its mean and variance.

We may simplify Eqs. 3.4 and 3.5 considerably if we are only interested in times much longer than the correlation time $\tau_c = m_p/\gamma$. The Stokes drag coefficient for a spherical particle of radius r is given by $\gamma = 6\pi r\eta$, where η is the viscosity of the solvent[41]. This implies that τ_c scales as r^2, and we compute $m_p/\gamma \approx 30$ns for a 1μm polystyrene sphere in water. With the exception of the DNA molecules discussed in Chapter 7, all of the particles that we study in this thesis are much smaller than 1μm so the relaxation of their velocity was too fast to be resolved by our apparatus. The apparatus could potentially resolve this time-scale on the DNA molecules, but not in the configuration that we used for making those measurements. As a result, the long-time approximation

$$\begin{aligned}\langle v_p(t)\rangle &= 0\\ \langle v_p(t_1)v_p(t_2)\rangle &= 2D\delta(t_2-t_1),\end{aligned} \tag{3.8}$$

where $D = k_BT/\gamma$ is the particle's *diffusion coefficient*, is always sufficient to describe the dynamics we observe in this thesis.

Given the random process v_p, all that remains is the integration

$$x_p(t) = \int_0^t \mathrm{d}\tau\, v_p(t) \tag{3.9}$$

from which we compute the statistical properties of x_p. Using Eq. 3.8, we get

$$\langle x_p(t)\rangle = 0 \tag{3.10}$$

$$\langle x_p(t_1)x_p(t_2)\rangle = 2D\min\{t_1,t_2\}, \tag{3.11}$$

which illustrates an important characteristic of the mean-squared displacement: it scales in proportion to t for Brownian motion, while it usually scales as t^2 for objects that move deterministically.

3.1.2 Localization noise process

The second input to the feedback loop is the localization noise $\mathbf{N}(t)$. Arising due to photon-counting noise as described in Section 2.2.4, $\mathbf{N}(t)$ is a zero-mean random process with variance determined by the beam geometry, averaging bandwidth B, electronic gain V_0A_V and photon counting rate Γ. Because the correlation time of the photon-counting fluctuations is very short (see Eq. 2.12), the correlation time of $\mathbf{N}(t)$ is set by the averaging as approximately B^{-1}. B is typically much larger than the tracking bandwidth, so B^{-1} is much shorter than the time-scales of the motion of the tracking stages. Therefore we may approximate the statistics

$$\langle \mathbf{N}(t)\rangle = 0 \tag{3.12}$$

$$\langle \mathbf{N}(t_1)\mathbf{N}(t_2)^T\rangle \approx n^2\,\mathrm{Id}\,\delta(t_1-t_2), \tag{3.13}$$

where n^2 is given by Eq. 2.17 or 2.28 and, by allowing the correlation matrix to be proportional to the identity, we have assumed that $\mathbf{N}(t)$ is not correlated along orthogonal axes. We may, as with $\mathbf{x}_p$, deal only with the scalar noise process $N(t)$ along any single axis. Because it is an average, we may approximate the distribution of $\mathbf{N}(t)$ as Gaussian by invoking the Central Limit theorem[8, e.g.], so that Eqs. 3.12 and 3.13 fully specify the process $\mathbf{N}(t)$.

3.1.3 Controller and plant dynamics

We assume that both the controller and the plant respond linearly to their inputs. While this may seem restrictive, most control electronics and many stage actuators — in particular, our piezoelectric actuators — are very nearly linear. Given this assumption, the inputs (generically denoted $u(t)$) and outputs ($y(t)$) of both blocks satisfy a dynamical system of the form

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{q}(t) &= \mathbf{A}\mathbf{q}(t) + \mathbf{B}u(t) \\ y(t) &= \mathbf{C}\mathbf{q}(t), \end{aligned} \tag{3.14}$$

where $\mathbf{q}$ is an internal state vector and the matrices $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$ are together referred to as the *state-space representation* of the system. An alternative representation of the system that is often useful is found by taking the Laplace transform of Eq. 3.14, giving

$$\tilde{y}(s) = \mathbf{C}(s\,\mathrm{Id} - \mathbf{A})^{-1}\mathbf{B}\tilde{u}(s). \tag{3.15}$$

The quotient $\tilde{y}(s)/\tilde{u}(s)$ is known as the *transfer function* from u to y. The notations $C(s)$ and $P(s)$ from Fig. 3.1 represent the transfer functions of the controller and plant, respectively. Their combined transfer function is simply the product $C(s)P(s)$ which, if not for the presence of the localization block, would be referred to as the *loop transfer function* because it maps the input to the output of the feedback loop.

3.1.4 Localization estimate

We let the localization estimation block be composed of a single function $L[\mathbf{e}]$ that maps $\mathbf{e}$ into $\langle\hat{\mathbf{e}}\rangle$. Then

$$\hat{\mathbf{e}}(t) = L[\mathbf{e}] + N(t). \tag{3.16}$$

In Sections 2.2.3 and 2.3.1 we calculated the exact expressions for L along the three axes of the localization coordinate system, showing that it is, in general, nonlinear. This nonlinearity can potentially make analysis quite difficult because nonlinear differential equations are hard to solve. Fortunately, $L[\mathbf{e}]$ is dominated by its linear component when $\mathbf{e}$ is small relative to the laser beam geometry. For closely-tracked particles we may write $L[\mathbf{e}] \approx \mathbf{e}$, provided that the electronic gain is set appropriately to calibrate the position estimate. In the next section, we derive the statistics of the localization error when L is approximated in this way.

3.2 Linearized tracking dynamics

If the localization estimate is a linear function of the tracking error, then all of the dynamics in the feedback loop are linear and we may study the statistics of the tracking error using ideas from linear stochastic control theory. This section is a review of Andy Berglund's work in this area[37, 42]. Those references are far more complete, as we only derive what we will need for the remainder of the work in this thesis. We take a different approach than those references, however, deriving all of the relevant statistics using the Langevin equation rather than the equivalent Fokker-Planck equation. In this section we deal only with the scalar variables x_p, x_s, e, $\hat{e}$ and N; as discussed in the previous section, the statistics of the corresponding vector variables are just vectors and diagonal matrices of the statistics of these scalars.

3.2.1 Input-output statistics

Given our input variables $x_p(t)$ and $N(t)$, the statistics of the remaining variables in the feedback loop are determined by input-output relationships that can be found by simple block-diagram algebra[43]. In terms of the controller and plant transfer functions, we have

$$\begin{aligned} \frac{\tilde{e}(s)}{\tilde{x}_p(s)} &= \frac{1}{1 + C(s)P(s)} \equiv T_1(s) \\ \frac{\tilde{e}(s)}{\tilde{N}(s)} = \frac{\tilde{x}_s(s)}{\tilde{x}_p(s)} = \frac{\tilde{x}_s(s)}{\tilde{N}(s)} &= \frac{C(s)P(s)}{1 + C(s)P(s)} \equiv T_2(s). \end{aligned} \tag{3.17}$$

All of these relationships have corresponding state-space realizations. Since our inputs were characterized in the time domain, we use the expressions in Eq. 3.17 as a means of determining the appropriate $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ for each relevant transfer function. We then express the statistics of the output signals in terms of these matrices.

The exact solution to the generic dynamical system in Eq. 3.14 for a particular realization of the random input process $u(t)$ is given by

$$y(t) = \mathbf{C}e^{\mathbf{A}t}\mathbf{q}_0 + \mathbf{C}e^{\mathbf{A}t}\int_0^t \mathrm{d}\tau\, e^{-\mathbf{A}\tau}\mathbf{B}u(\tau). \tag{3.18}$$

We are interested not in particular realizations, but rather in the statistics of the random process $y(t)$. We may compute the moments directly from Eq. 3.18 in the same manner as we did in Section 3.1.1:

$$\langle y(t) \rangle = \mathbf{C}e^{\mathbf{A}t}\mathbf{q}_0 \tag{3.19}$$

$$\langle\langle y(t_1)y(t_2)\rangle\rangle = \mathbf{C}\int_0^{t_1} \mathrm{d}\tau_1 \int_0^{t_2} \mathrm{d}\tau_2\, e^{\mathbf{A}(t_1-\tau_1)}\mathbf{B}\langle\langle u(\tau_1)u(\tau_2)\rangle\rangle\mathbf{B}^T e^{\mathbf{A}^T(t_2-\tau_2)}\mathbf{C}^T. \tag{3.20}$$

Special consideration must be given to the two cases $u(t) = N(t)$ and $u(t) = x_p(t)$ due to fundamental differences arising from the forms of their correlation functions (Eqs. 3.11 and 3.13). In particular, $N(t)$ is a delta-correlated stationary process and $x_p(t)$ is not, and this difference has a major effect on the statistics of the output signals.

3.2.1.1 The case $u(t) = N(t)$

Since $N(t)$ is delta-correlated, the integrals in Eq. 3.20 collapse into the single integral

$$\langle\langle y_N(t_1) y_N(t_2)\rangle\rangle = n^2 \mathbf{C} \int_0^{\min\{t_1,t_2\}} \mathrm{d}\tau \, e^{\mathbf{A}(t_1-\tau)} \mathbf{B}\mathbf{B}^T e^{\mathbf{A}^T(t_2-\tau)} \mathbf{C}^T, \tag{3.21}$$

where the subscript notation is used to keep track of the input signal. The explicit dependence of this expression on t_1 and t_2 is a consequence of our derivation of it by solving an initial value problem. This is clear from the fact that $\langle y(0)\rangle = \mathbf{C}\mathbf{q}_0$ and $\langle\langle y(0)^2\rangle\rangle = 0$. This transient artifact is eliminated by setting $t_2 = t_1 + \tau$ and taking the limit $t_1 \to \infty$. The resulting integral

$$\Sigma^\infty \equiv \lim_{t_1\to\infty} \int_0^{t_1} \mathrm{d}\tau \, e^{\mathbf{A}(t_1-\tau)} \mathbf{B}\mathbf{B}^T e^{\mathbf{A}(t_1-\tau)} \tag{3.22}$$

converges when the eigenvalues of $\mathbf{A}$ are all negative, which is a prerequisite for stability of the feedback system[43]. We are left with the simple expression

$$\langle\langle y_N(t+\tau) y_N(t)\rangle\rangle = n^2 \mathbf{C} e^{\mathbf{A}\tau} \Sigma^\infty \mathbf{C}^T. \tag{3.23}$$

3.2.1.2 The case $u(t) = x_p(t)$

Things get complicated when $u(t) = x_p(t)$ because x_p is not delta-correlated. Neither integral in Eq. 3.20 disappears, and the minimum function in Eq. 3.11 requires us to divide the integration region into three parts based on the relationship between t_1 and t_2 in each part. The integral does not converge as $t \to \infty$, making matters even worse.

The calculation is simplified somewhat if we consider the transfer function from $v_p(t)$, rather than $x_p(t)$, to $y(t)$. The integral relating $v_p(t)$ to $x_p(t)$ gives us

$$\frac{\tilde{y}(s)}{\tilde{v}_p(s)} = \frac{1}{s} \frac{\tilde{y}(s)}{\tilde{x}_p(s)}, \tag{3.24}$$

which says that using $v_p(t)$ as our input to the system produces precisely the time derivative $\dot{y}(t) = \mathrm{d}y(t)/\mathrm{d}t$. The statistics of $\dot{y}(t)$ are much simpler than those of $y(t)$ because $v_p(t)$ is delta-correlated. In fact, they are given exactly by Eq. 3.23, with $2D$ in place of n^2. We may now compute the statistics of $y(t)$ by integrating a much simpler expression; we get

$$\begin{aligned}\langle\langle y_{x_p}(t+\tau)y_{x_p}(t)\rangle\rangle &= \int_0^{t+\tau}\mathrm{d}t_1\int_0^{t}\mathrm{d}t_2\langle\langle\dot{y}(t_1)\dot{y}(t_2)\rangle\rangle\\ &= 2D\mathbf{C}\mathbf{A}^{-2}\left(e^{\mathbf{A}t}+e^{\mathbf{A}(t+\tau)}-e^{\mathbf{A}\tau}-2\mathbf{A}t-\mathrm{Id}\right)\Sigma^{\infty}\mathbf{C}^{T}. \qquad (3.25)\end{aligned}$$

3.2.2 Second-order dynamics

So far, the expressions we have derived for the input-output statistics of the feedback loop allow for the controller and plant transfer functions to be chosen arbitrarily. This flexibility provides a general tool for describing the dynamics of particle-tracking systems, but in our case we are interested in only one particular controller and plant pair.

We use a purely integrating feedback controller, $C(s) = \gamma_c/s$. We choose this primarily because of its simplicity, but it can also be shown that it is optimal in the sense that it minimizes the least-squared tracking error for a fixed localization noise density and a plant with a flat frequency response[37, 42]. The factor of γ_c is a gain term; actually, γ_c is the entire loop gain of the tracking system and is the product of the fluorescence rate, lock-in amplifier gain, integrator gain, and the voltage-to-position conversion factor of the tracking stages. In this sense γ_c is the only adjustable parameter in the model; we typically adjust it by varying the excitation laser power (and consequently the fluorescence rate) and the lock-in amplifier gain.

The plant dynamics are generally fairly complicated, with multiple resonances that are difficult to model accurately. However, the amplifier that drives the stages has a built-in low-pass filter, and the choice $P(s) = \gamma_p/(s+\gamma_p)$, where γ_p is the roll-off frequency of the filter, provides a rich enough model to describe the dynamics we usually observe[37, 42].

Given these choices for $C(s)$ and $P(s)$, a valid state-space representation for the transfer function $T_2(s)$ from Eq. 3.17 is

$$T_2(s) = \frac{\gamma_c\gamma_p}{s^2+\gamma_p s+\gamma_c\gamma_p} \Rightarrow \mathbf{A}_2 = \begin{pmatrix} -\gamma_p & -\gamma_p \\ \gamma_c & 0 \end{pmatrix}, \mathbf{B}_2 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \mathbf{C}_2 = \begin{pmatrix} 0 & \gamma_p \end{pmatrix}. \qquad (3.26)$$

$T_2(s)$ is stable because it is nonzero at all finite values of s[44]. This fact is further supported by the fact that $\mathbf{A}_2$ has no nonnegative eigenvalues; this means that Σ^{∞} as defined in Eq. 3.22 does indeed converge.

In contrast, $T_1(s)$ is not stable because it has a zero at $s = 0$[44]. This fact applies generally for any integrating controller, since an integrator has infinite gain at $s = 0$. We are fortunate enough to be able to avoid any difficulties arising from this instability through a slight manipulation of $T_1(s)$. Since $T_1(s)$ is a transfer function from x_p only, we may convert it to a transfer function from v_p by incorporating the integral relating these two signals into the transfer function rather than the input. We get

$$\frac{1}{s}T_1(s) = \frac{s+\gamma_p}{s^2+s\gamma_p+\gamma_c\gamma_p} \Rightarrow \bar{\mathbf{A}}_1 = \mathbf{A}_2, \bar{\mathbf{B}}_1 = \mathbf{B}_2, \bar{\mathbf{C}}_1 = \begin{pmatrix} 1 & \frac{\gamma_p}{\gamma_c} \end{pmatrix}, \tag{3.27}$$

where the ¯ indicates that this realization corresponds to the modified transfer function $T_1(s)/s$, *not* $T_1(s)$. The integral canceled the zero at $s = 0$ exactly, giving a stable relationship between $v_p(t)$ and the output $e(t)$. Such cancellation is sometimes risky because it implies that the feedback system is not *internally* stable[44] — a very large signal is used to cancel a very small signal, but very large signals often cause problems. For example, in electronics this cancellation would require that the large signal is not allowed to saturate, but that is never the case. In our tracking application the internal instability corresponds to the fact that $x_p(t)$ is allowed to get very big because we do not limit the particle's motion. The only practical problem with this is that $x_p(t)$ will eventually exceed the limits imposed by our tracking stages, so that the tracking system fails to continue tracking the particle.

We must be sure to calculate the output statistics for the modified system $T_1(s)/s$ using Eq. 3.23, with $2D$ in place of n^2. We do not use Eq. 3.25 because the integration of v_p is already incorporated into the realization. For both $T_2(s)$ and $T_1(s)/s$, we compute the same expression for Σ^∞:

$$\Sigma^\infty = \begin{pmatrix} \frac{1}{2\gamma_p} & 0 \\ 0 & \frac{\gamma_c}{2\gamma_p^2} \end{pmatrix}. \tag{3.28}$$

3.2.2.1 Tracking error statistics

For our work in the next chapter we will compute the contribution to the fluorescence signal made by the statistics of the tracking error. To do this we require the covariance matrix of the joint probability density $p(\mathbf{e}^{t+\tau}, \mathbf{e}^t)$. We use Eq. 3.23 to compute the elements of this matrix. Defining $\sigma_\alpha^\tau = \left[\langle e_{x_p}(t+\tau)e_{x_p}(t)\rangle + \langle e_N(t+\tau)e_N(t)\rangle\right]^{1/2}$, where the subscript $\alpha \in \{x, y, z\}$ denotes a particular Cartesian axis, we have

$$\sigma_\alpha^\tau = \sqrt{2D\bar{\mathbf{C}}_1 e^{\bar{\mathbf{A}}_1\tau}\Sigma^\infty\bar{\mathbf{C}}_1^T + n^2\mathbf{C}_2 e^{\mathbf{A}_2\tau}\Sigma_\infty\mathbf{C}_2^T}. \tag{3.29}$$

This expresses the correlation function of the total tracking error, arising from both the particle's motion and the localization noise. With this definition, we write the covariance matrix for $p(\mathbf{e}_\alpha^{t+\tau}, \mathbf{e}_\alpha^t)$ as

$$\Sigma_{\mathbf{e}_\alpha}^\tau = \begin{bmatrix} (\sigma_\alpha^0)^2 & (\sigma_\alpha^\tau)^2 \\ (\sigma_\alpha^\tau)^2 & (\sigma_\alpha^0)^2 \end{bmatrix}. \tag{3.30}$$

Since there is no correlation of the error between different Cartesian axes, the covariance matrix $\Sigma_{\mathbf{e}}^{\tau}$ for the joint vector probability density $p(\mathbf{e}^{t+\tau}, \mathbf{e}^{t})$ is block-diagonal with $\Sigma_{\mathbf{e}_x}$, $\Sigma_{\mathbf{e}_y}$ and $\Sigma_{\mathbf{e}_z}$ along the diagonal.

The symbolic expression in Eq. 3.29 can be evaluated by computing the matrix exponentials; the result is given in [37, 42] and is fairly complicated, so we do not include it here. One interesting quantity that is not overly complicated is that for the steady-state root-mean-squared tracking error:

$$\sigma_{\alpha}^{0} = \sqrt{D\left(\frac{1}{\gamma_c} + \frac{1}{\gamma_p}\right) + \frac{n^2\gamma_c}{2}}. \tag{3.31}$$

This expression provides insight into how the parameters of the model, localization noise and Brownian motion affect the error statistics. As discussed above, γ_c is easily adjustable, so we consider what happens when we vary it. Increasing γ_c reduces the first term under the radicand because it improves the ability to track the particle's motion, but simultaneously increases the size of the error induced by the localization noise. Even for very small n, however, increasing γ_c is only effective to an extent. Returns diminish when γ_c compares to or exceeds γ_p because at this point the response of the tracking stage limits tracking accuracy, while the localization noise fed into the tracking error continues to grow. The optimal value occurs at $\gamma_c = \sqrt{2D/n^2}$, independent of γ_p, giving

$$\min_{\gamma_c} \sigma_{\alpha}^{0} = \sqrt{\sqrt{2n^2D} + \frac{D}{\gamma_p}}. \tag{3.32}$$

This result illustrates the existence of a bandwidth-limited tracking regime — even in the absence of localization noise, nothing within our control can improve the localization to better than $\sqrt{D/\gamma_p}$. This has serious implications, requiring consideration of the nonlinearity in the localization estimate if this optimal error is too large. We discuss this more in Section 3.3.

3.2.2.2 Stage position increment statistics

A final topic we have yet to discuss is the determination of the parameters in our model. While we can measure certain parameters directly — γ_p is easy to measure, and γ_c is not too difficult — we require a systematic approach to the estimation of the diffusion coefficient and localization noise density. One approach uses fluctuations in the fluorescence signal induced by the tracking errors we just discussed; that approach is presented in Chapter 4. The accuracy with which this can be done depends on the size of the fluctuations, however, and the fluctuations become smaller as the particle is tracked better. On the z axis in particular, it is often impossible to infer the tracking statistics from the fluorescence signal.

Our tracking stages contain capacitive sensors that allow us to monitor their positions over time. We can therefore directly measure the statistics of $\mathbf{x}_s$ and use these to infer the model

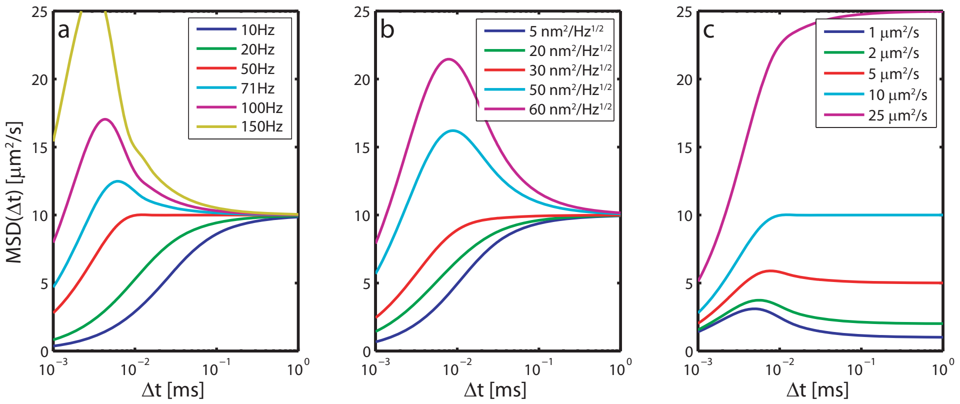


Figure 3.4: $MSD(\Delta t)$ with varied second-order system parameters. **a**: γ_c varied with fixed $\gamma_p = 100\text{Hz}$, $D = 10\mu\text{m}^2/\text{s}$ and $n = 0.01\mu\text{m}/\sqrt{\text{Hz}}$. **b**: n varied with fixed $\gamma_c = 20\text{Hz}$, $\gamma_p = 100\text{Hz}$, $D = 10\mu\text{m}^2/\text{s}$. **c**: D varied with fixed $\gamma_c = 50\text{Hz}$, $\gamma_p = 100\text{Hz}$, $D = 10\mu\text{m}^2/\text{s}$ and $n = 0.01\mu\text{m}/\sqrt{\text{Hz}}$.

parameters. We compute the mean-squared deviation of the stage position increment as a function of the increment time Δt,

$$MSD(\Delta t) \equiv \frac{1}{2\Delta t}\left\langle [x_s(t+\Delta t) - x_s(t)]^2 \right\rangle, \tag{3.33}$$

which is defined in this way so that $\lim_{\Delta t\to\infty} MSD(\Delta t) = D$. As with the tracking error, $MSD(\Delta t)$ is found by adding the contributions to it from both the particle's motion and the localization noise. In terms of the state-space realization for $T_2(s)$, we have

$$MSD(\Delta t) = \frac{2D}{\Delta t}\mathbf{C}_2\mathbf{A}_2^{-2}\left[e^{\mathbf{A}_2\Delta t} - \mathbf{A}_2\Delta t - \text{Id}\right]\Sigma^\infty\mathbf{C}_2^T + \frac{n^2}{\Delta t}\mathbf{C}_2\left[\text{Id} - e^{\mathbf{A}_2\Delta t}\right]\Sigma^\infty\mathbf{C}_2^T. \tag{3.34}$$

This symbolic expression can also be written in terms of the individual parameters γ_c and γ_p by directly computing the matrix exponentials. Unlike Eq. 3.29, this less abstract form is not given in other references. It is particularly useful because we often compute numerical fits to $MSD(\Delta t)$, so efficient calculation is important. We give the simplified expression here, because it would be a shame for a third person (after Andy Berglund and, now, me) to have to compute this complicated quantity:

$$\begin{aligned} MSD(\Delta t) = D - \frac{D}{\Delta t}e^{-\gamma_p\Delta t/2}\Bigg\{\frac{2}{\nu}\sinh\left(\frac{\nu\Delta t}{2}\right) + \left(\frac{1}{\gamma_p} - \frac{1}{\gamma_c} + \frac{\gamma_c n^2}{2D}\right) \\ \times\left[\cosh\left(\frac{\nu\Delta t}{2}\right) + \frac{\gamma_p}{\nu}\sinh\left(\frac{\nu\Delta t}{2}\right) - e^{\gamma_p\Delta t/2}\right]\Bigg\}, \end{aligned} \tag{3.35}$$

where we have defined $\nu = \sqrt{\gamma_p^2 - 4\gamma_c\gamma_p}$.

Figure 3.4 illustrates the $MSD(\Delta t)$ curves predicted by Eq. 3.35 for several sets of second-order system parameters. All curves share $MSD(0) = 0$ because the tracking stage moves with a finite bandwidth and $MSD(\infty) = D$ because the dynamics of the tracking stage become identical to those of the particle at long times. The rise time and overshoot, and more specifically the functional form of $MSD(\Delta t)$ on intermediate time-scales, are consequences of the system parameters. Plot **a** shows that the rise times decrease with increasing loop gain but that there is significant overshoot when γ_c is too large due to the phase lag in the plant dynamics. The curve at $\gamma_c = 71\text{Hz}$ represents the optimal value for this parameter. Plot **b** shows that localization noise can have a dramatic effect on the stage dynamics at short time-scales, but must be suppressed on longer time-scales because otherwise the tracking system would not follow the particle. Plot **c** shows that the effect of $N(t)$ on $MSD(\Delta t)$ depends in part on D: it is more noticeable when D is small compared to n.

3.3 Nonlinear limitations in real tracking systems

The previous section illustrated the importance of the simplification of the tracking system dynamics resulting from setting $L[\mathbf{e}] = \mathbf{e}$. We were able to fully characterize the statistics of all relevant variables in the feedback loop. In doing this we showed the existence of a bandwidth-limited tracking regime (see Eq. 3.32) in which our localization statistics cannot be improved by adjusting the gain in the feedback loop and depend on the diffusion coefficient D. This means that for any real tracking system with finite actuation bandwidth γ_p, the linearity approximation is guaranteed to fail for sufficiently small particles (with sufficiently large D). In this section, we consider this fundamental limit to tracking system performance.

The standard deviation of the tracking error under the linear approximation (with optimal γ_c) is given by Eq. 3.32. As shown in Section 2.2.3, the linear localization region in two dimensions is the disk bounded by the rotation radius r, related to the beam waist w_0 by $w_0 \approx r\sqrt{2}$. This allows us to calculate the diffusion coefficient threshold for remaining within the linear regime with 95% probability:

$$D \lesssim n^2\gamma_p^2 + \frac{w_0^2\gamma_p}{8} - \frac{n\gamma_p^2}{2}\sqrt{4n^2 + \frac{w_0^2}{\gamma_p}}. \tag{3.36}$$

It is important to remember that this provides an overestimate of the actual threshold; any calculations that use the linear approximation to compute properties of the outer bounds of that region will be influenced by the breakdown of the approximation near those bounds. The nonlinear terms detract from the localization estimate ($|L[e]| \le |e|$), and this implies that the values we calculate are always optimistic in terms of the size of the tracking error. Despite this, it may be argued that the ease with which such calculations are made offsets their imprecision.

When D is greater than the threshold in Eq. 3.36, the tracked particle spends some of its time at the fringes of the localization signal. The linearity approximation breaks down, making it very difficult to analytically compute statistics such as σ_α^τ and $MSD(\Delta t)$. In fact, since $L[e]$ decays to zero for large e the particle will, with probability 1, eventually escape the tracking apparatus so that our entire model of the tracking system breaks down. We take an approach in this section that is meant to characterize this phenomenon — to determine some simple statistics of escape.

Throughout the remainder of this section we set $n = 0$ — this just means that we are capable of collecting enough photons from the particle so that localization noise does not limit our tracking fidelity. Our concern is therefore strictly with tracking inaccuracy due to the particle's movement, although it is just a generalization of our approach here to incorporate localization noise into our model. For a complete discussion of localization noise-limited tracking see [36, 37].

3.3.1 Bandwidth-limited steady-state error

We may evaluate just how the nonlinear position estimate affects the tracking statistics by considering the steady-state distribution of the particle relative to tracking stage. For simplicity, we will approximate $P(s) = 1$, so that the tracking system has first-order dynamics. While this model does not actually exhibit bandwidth-limited feedback because taking $\gamma_c \to \infty$ (with $n = 0$) reduces the tracking error to zero, any finite γ_c yields the tracking error $\sigma_\alpha^0 = \sqrt{D/\gamma_c}$ and the diffusion coefficient threshold $D \lesssim w_0^2\gamma_c/8$.

We may write a stochastic differential equation for the error:

$$\frac{\mathrm{d}}{\mathrm{d}t}e(t) = v_p(t) - \gamma_c L[e(t)]. \tag{3.37}$$

This nonlinear equation is difficult to solve exactly, but it can be shown that it is equivalent to the Fokker-Planck equation for the probability density function $p(e,t)$[8],

$$\frac{\partial}{\partial t}p = \gamma_c \frac{\partial}{\partial e}\left(L[e]p\right) + D\frac{\partial^2}{\partial e^2}p, \tag{3.38}$$

which we can use relatively easily to analyze the statistics of e because it is a linear partial differential equation. We know that there is a nonzero probability that the particle will escape from the tracking system because L decays to zero for large tracking errors. As an extreme example, as $\gamma_c \to 0$ the particle is not tracked at all and the statistics of the error are those of pure, unconstrained Brownian motion. Therefore, $e(t)$ is not a stationary process unless we impose boundary conditions on it; for the sake of analysis, we introduce the arbitrary bound

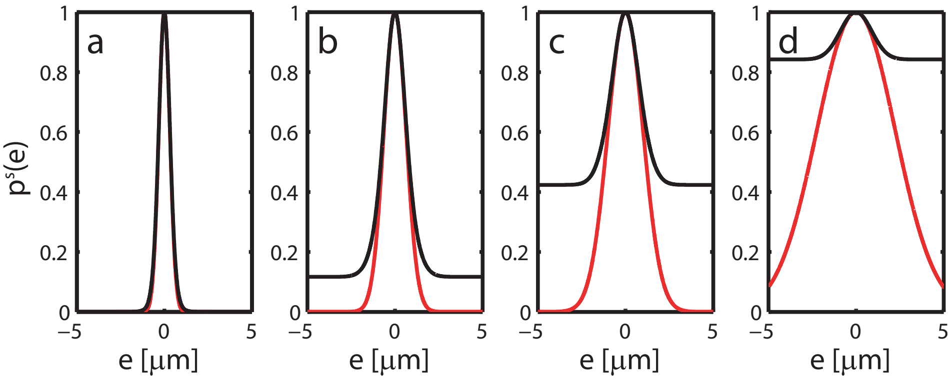


Figure 3.5: Steady-state tracking error due to nonlinear $L[e]$ (black) compared to that predicted by the linearized theory (red). $\gamma_c/D = \{$**a**: 10; **b**: 2.5; **c**: 1; **d**: 0.2$\}$ [μm^{-2}]. Both curves in each plot are normalized so that their peaks have value 1. In all plots we used the calibrated $L[e]$ for radial localization (Eqs. 2.13 and 2.15) with $r = 1\mu$m and $w_0 = r\sqrt{2}$.

$-a < e < a$, so that

$$\left.\frac{\partial}{\partial e} p\right|_{e=\pm a} = 0. \tag{3.39}$$

Solving this boundary value problem exactly would require a series solution. However, we find that $L[e]$ nearly imposes the boundary condition itself for large enough a: solving Eq. 3.38 without regard for the boundary condition gives

$$p^s(e) \propto \exp\left(-\frac{\gamma_c}{D}\int_0^e \mathrm{d}e' L[e']\right), \tag{3.40}$$

and the derivative $\mathrm{d}p^s(e)/\mathrm{d}e$ contains a multiplicative factor of $L[e]$. If we choose a so that $L[a] \approx 0$, then this $p^s(e)$ approximately solves the boundary value problem.

Figure 3.5 shows the stationary distribution $p^s(e)$ for a particle confined so that $|e| < 5\mu$m. The flatness of the (nonlinear) curves at the boundaries indicates that the boundary conditions in Eq. 3.39 are approximately satisfied. We find close agreement between the predictions of the nonlinear and linearized theory for large γ_c/D, but the theory breaks down dramatically when the particle is tracked poorly. The nonlinear model predicts that confinement only occurs near the linear region $|e| < r$, while the linear model predicts that the width of the error distribution may increase arbitrarily as γ_c/D decreases. The failure of the tracking system in the badly-localized cases is reflected by the increased probability density outside the linear region.

Figure 3.6 shows the steady-state probability integrated over the linear region as a function of γ_c/D. At small γ_c/D, the linear region's occupancy is roughly equal to the value 0.2, corresponding to a completely unconfined particle. A sharp transition occurs between about 1 and $10\mu m^{-2}$: for $\gamma_c/D > 10\mu$m^{-2}, the particle almost never exits the linear region. This

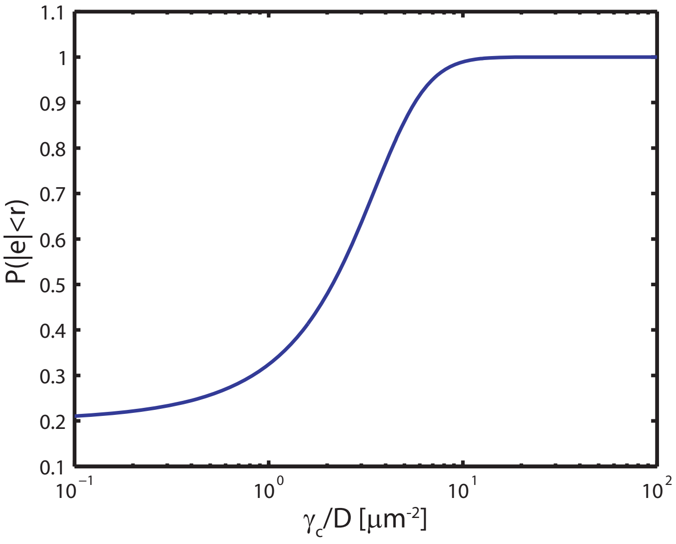


Figure 3.6: Steady-state occupancy of the linear region. We computed Eq. 3.40 at varied values of γ_c/D and integrated the stationary distribution over the linear region $|e| < r$. We fixed $w_0 = r\sqrt{2}$ in these calculations.

abrupt transition suggests that there is not much room for a “nonlinear tracking” regime — the particle is apparently either tracked well, with a stationary distribution that closely resembles the predictions of the linearized theory, or it is not tracked at all. We consider this idea more thoroughly in the next section.

3.3.2 Escape time statistics

As shown in the last section, the effect of the nonlinearity in $L[e]$ on the steady-state distribution of the tracking error is fairly dramatic. There is a sharp transition from apparently not tracking at all to tracking perfectly over an order of magnitude increase in the tracking bandwidth. However, it is not obvious how to interpret this in a dynamic context; a given steady-state distribution tells us nothing about the rate at which the particle transitions from being tracked to not being tracked. Such dynamic properties are most appropriately studied using the statistics of first passage times[8, 45], in which we treat the amount of time that a particular particle is tracked for as a random variable and compute the statistics of that variable.

A slight ambiguity exists in our description of the duration of a tracking trajectory: there is no obvious way to determine precisely when tracking begins or ends, because the transition between the two is very smooth. The untracked particle simply drifts away from the tracking stage; it is not pulled away or frozen in place or changed in any way that would provide a good criterion for being lost by the apparatus. Our only option is to choose a threshold value ε and posit that the apparatus loses the tracked particle if $|e|$ reaches or exceeds ε. While

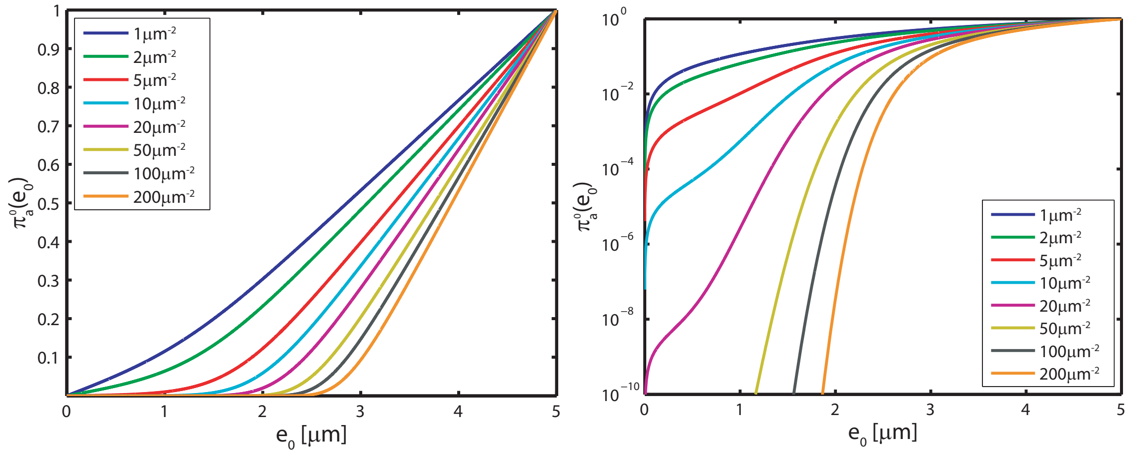


Figure 3.7: Splitting probability $\pi_a^0(e_0)$ from Eq. 3.42. The plot on the left illustrates the clear transition between the tracked and untracked regions. The plot on the right shows the same curves on a log scale to highlight the extremely small exit probability within the tracked region. We varied γ_c/D as indicated in the legend and fixed $r = 1\mu$m and $w_0 = \sqrt{r}$. For $L[e]$ we used the calibrated radial localization signal from Eqs. 2.13 and 2.15.

setting thresholds is usually an ambiguous process itself, we can approach this problem in a systematic way and find a solution that yields consistent threshold values.

We again require that the particle reside in a box of length a, so that $|e| < a$. We define the *splitting probability* $\pi_a^0(e_0)$ representing the probability that the tracking error, initially $e = e_0$, reaches $e = a$ before reaching $e = 0$. We may likewise define $\pi_0^a(e_0) = 1 - \pi_a^0(e_0)$. It may be shown that $\pi_a^0(e_0)$ obeys the adjoint differential equation to that obeyed by the steady-state distribution $p^s(e)$[8]. This gives

$$-\gamma_c L[e_0]\frac{\mathrm{d}}{\mathrm{d}e_0}\pi_a^0(e_0) + D\frac{\mathrm{d}^2}{\mathrm{d}e_0^2}\pi_a^0(e_0) = 0 \tag{3.41}$$

with the boundary conditions $\pi_a^0(a) = 1$, $\pi_a^0(0) = 0$. This boundary value problem is solved exactly by

$$\pi_a^0(e_0) \propto \int_0^{e_0} \mathrm{d}e' \, \exp\left(\frac{\gamma_c}{D}\int_0^{e'} \mathrm{d}e'' \, L[e'']\right), \tag{3.42}$$

where the proportionality constant is found simply by enforcing the boundary condition at $e_0 = a$.

Figure 3.7 shows $\pi_a^0(e_0)$ for various values of γ_c/D with fixed r and w_0. All of the curves show a clear transition between a tracked regime — in which the exit probability π_a^0 was extraordinarily small — to an untracked regime in which the exit probability became an approximately linear function of position. In the absence of any feedback ($L[e] = 0$), the splitting probability is exactly linear; the linearity we observe therefore implies that in the untracked region the particle's motion is not much different than it would be if the particle were freely diffusing.

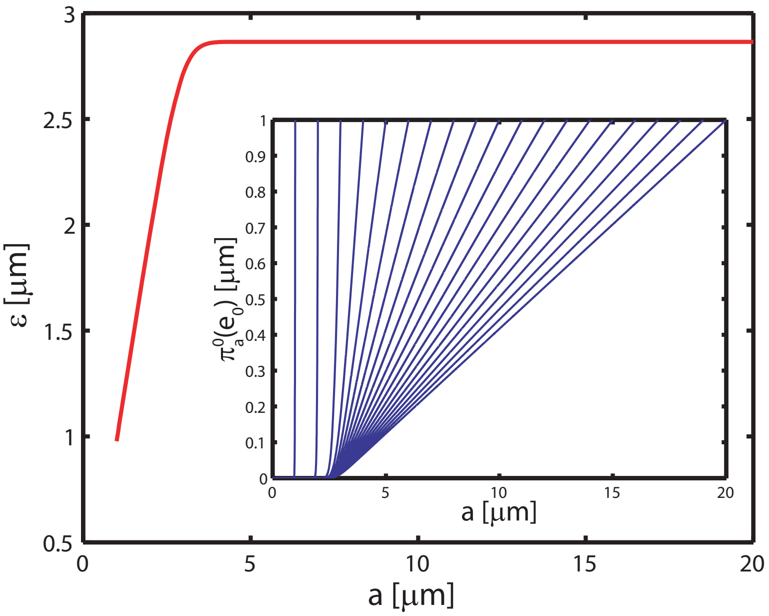


Figure 3.8: Choosing ε by the extrapolation procedure described in the text. We varied a and computed ε for fixed $\gamma_c/D = 50\mu\text{m}^{-2}$ by extrapolating the linear part of $\pi_a^0(e_0)$ to zero. Inset shows $\pi_a^0(e_0)$ resulting from variations in a. Both plots used $r = 1\mu\text{m}$, $w = r\sqrt{2}$.

There is one caveat to choosing ε based on the splitting probability that we must consider: the choice of a is arbitrary, but $\pi_a^0(e_0)$ depends strongly on it. For example, we could choose to assign ε based on an absolute value for the splitting probability — say, the point at which $\pi_a^0(\varepsilon) = 1\%$. The choice of a biases this approach because ε moves farther from 0 as a increases simply because it is less likely for the particle to first reach a more distant point. Instead, we choose ε by computing a linear fit $\hat{\pi}_a^0(e_0)$ to $\pi_a^0(e_0)$ near $e_0 = a$ and extrapolating the fit to satisfy $\hat{\pi}_a^0(\varepsilon) = 0$.

Figure 3.8 shows the choice of ε as a is varied. Once a is sufficiently large, ε chosen by extrapolation does not depend at all on a. It is important to remember that the choice *does* depend on the ratio γ_c/D; for example, a fast-moving particle is more likely to escape when it reaches a particular e_0 than a slow-moving particle is at the same e_0. Interestingly, our extrapolation method produces values of ε that are larger than r by almost a factor of 3. These values will of course depend on γ_c, but nonetheless suggest that the tracking system is capable of significantly suppressing tracking errors well beyond the linear localization regime.

Now that we have a systematic means of defining when a particle escapes the tracking apparatus, we may compute the statistics of escape events. We bound the tracking error by $|e| < \varepsilon$, and compute the mean time required for the error to first reach $\pm\varepsilon$ if it starts at a particular e_0. We denote the probability density function for the particle's escape time, depending on the

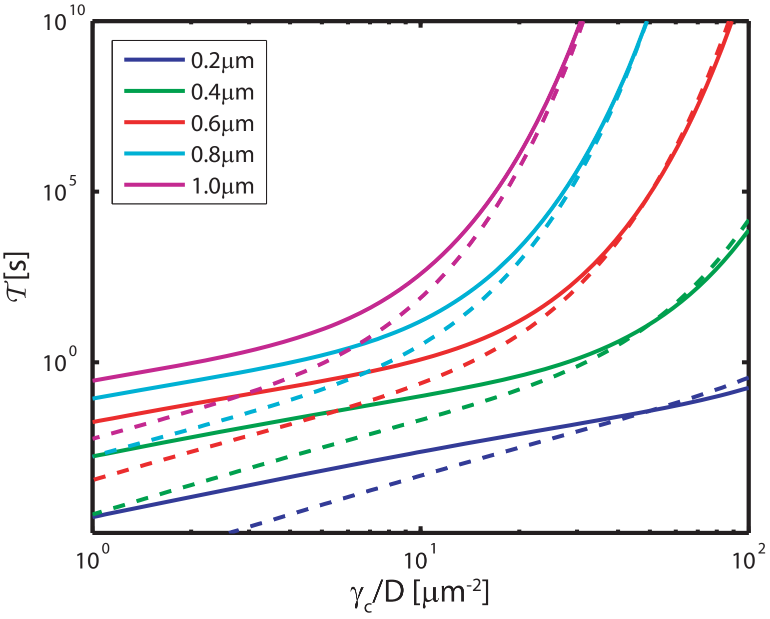


Figure 3.9: Mean escape time of a tracked particle as a function of γ_c/D for varied rotation radius r, indicated in the legend. We fixed $D = 1\mu\text{m}^2/\text{s}$ (solid curves) and $\gamma_c = 50\text{rad/s}$ (dashed curves). All curves used $w_0 = r\sqrt{2}$.

initial error e_0, by $p^{\varepsilon}_{e_0}(\tau)$. Since the particle escapes at either $e = \varepsilon$ or $e = -\varepsilon$, we may write

$$p^{\varepsilon}_{e_0}(\tau) = p(\tau|\varepsilon)\pi^{-\varepsilon}_{\varepsilon}(\varepsilon) + p(\tau|-\varepsilon)\pi^{\varepsilon}_{-\varepsilon}(e_0), \tag{3.43}$$

where the conditional probability notation indicates which boundary the particle exited from and the splitting probabilities are defined as before but for the boundaries $\pm\varepsilon$. We define the conditional mean first passage time

$$\vartheta^{-\varepsilon}_{\varepsilon}(e_0) = \int_0^{\infty} \mathrm{d}\tau\, \tau\, p(\tau|\varepsilon)\pi^{-\varepsilon}_{\varepsilon}(e_0). \tag{3.44}$$

It can be shown that $\vartheta^{-\varepsilon}_{\varepsilon}(e_0)$ also satisfies a differential equation adjoint to that for $p^s(e)$[8]:

$$-\gamma_c L[e_0]\frac{\mathrm{d}}{\mathrm{d}e_0}\vartheta^{-\varepsilon}_{\varepsilon}(e_0) + D\frac{\mathrm{d}^2}{\mathrm{d}e_0^2}\vartheta^{-\varepsilon}_{\varepsilon}(e_0) = -\pi^{-\varepsilon}_{\varepsilon}(e_0) \tag{3.45}$$

with the boundary conditions $\vartheta^{-\varepsilon}_{\varepsilon}(-\varepsilon) = 0 = \vartheta^{-\varepsilon}_{\varepsilon}(\varepsilon)$. The solution to this boundary value problem can be found explicitly just as it was for $\pi^0_a(e_0)$ in Eq. 3.42. Its form, however, is fairly complicated and given in [8], so we do not provide it here. Due to the symmetry of the localization signal, we know $\vartheta^{-\varepsilon}_{\varepsilon}(e_0) = \vartheta^{\varepsilon}_{-\varepsilon}(-e_0)$. Therefore, we need to solve for just one of these functions. We define the mean escape time $\mathcal{T}$ as the time it takes for a particle, beginning at the origin, to escape through either boundary: $\mathcal{T} = 2\vartheta^{-\varepsilon}_{\varepsilon}(0)$.

Figure 3.9 shows $\mathcal{T}$ as a function of γ_c/D, for varied rotation radii. Due to the form of Eq. 3.45, $\mathcal{T}$ depends explicitly on D (not just on γ_c/D), so we generated the curves in the plot both by varying γ_c with fixed D and by varying D with fixed γ_c. The curves suggest that the explicit dependence of $\mathcal{T}$ on D is most significant for poorly-tracked particles (with small γ_c/D). All curves show $\mathcal{T}$ increasing linearly on the log-log plot at small γ_c/D with a slope corresponding to a $\mathcal{T} \sim (\gamma_c/D)^2$ scaling as is predicted for unconstrained Brownian motion (see Eq. 3.11). For larger γ_c/D, a transition to exponential scaling of $\mathcal{T}$ occurs. The transition is extremely abrupt, with $\mathcal{T}$ increasing from 1s to over 10^5s with less than an order of magnitude increase in γ_c/D for most sets of parameters. This suggests that there is only a very small intermediate range of γ_c/D between regimes of completely failed tracking and nearly escape-proof tracking.

Our results in this section indicate that the most essential consequence of the nonlinearity of $L[e]$ is to impose a hard upper limit on the diffusion coefficient of particles that a particular feedback system is capable of tracking. While this conclusion was drawn from a fairly artificial model in which our own choice of γ_c was the only limitation to the tracking system's performance, it should apply quite generally to real systems with limitations arising from the imperfect frequency response of the plant or from localization noise.

The abruptness of the transition between trackable and untrackable diffusion coefficients is remarkable. At the same time, it is somewhat troubling: an apparatus that tracks 20nm particles with ease, for instance, might fail miserably on 5nm particles and it can be difficult to predict this. Such a failure could mean the difference between doing cutting-edge work on protein dynamics and being relegated to working with polystyrene beads. Fortunately, there are ways around some of these tracking limitations. The relatively slow $\sqrt{t}$ scaling of the mean-squared displacement of a Brownian particle means that great gains in trackability can be made just by increasing the size of the tracking laser beams, as is evident from the plots in Fig. 3.9. Similarly, this scaling implies that trackability improvements are possible by introducing a method for fast actuation of the stage position even if that actuation has a very limited range. In our apparatus, this could be done by using our acousto-optics (discussed in Chapter 5) to track small displacements of the particle while larger displacements are handled by the piezoelectric stages. The results in this section therefore largely apply to problems that can be eliminated by good design. Nonetheless, understanding these technical challenges is an important part of the further development of closed-loop particle-tracking systems.

Chapter 4

Fluorescence correlation spectroscopy

Having discussed our procedure for localizing fluorescent particles in Chapter 2 and the dynamics of closed-loop tracking systems in Chapter 3, we fully understand how closed-loop particle tracking works. However, we have not yet discussed the interpretation of the data that we collect from tracking experiments. In this chapter we discuss Fluorescence Correlation Spectroscopy (FCS), a statistical method based on the two-time correlation $g_2(\tau)$[8] of the fluorescence rate. This popular technique typically refers to experiments, such as those described in Chapter 1, in which fluorescence fluctuations are measured on small numbers of particles that are free to move in solution but are *not* tracked in closed-loop; however, the generalization to closed-loop FCS is straightforward, and this technique is a powerful approach to the analysis the fluorescence signals that we collect while tracking.

The development of FCS began at Cornell in 1972 when Magde, Elson and Webb studied the binding and unbinding of ethidium bromide to DNA[1]. Using an Ar^+ laser to excite and a parabolic reflector to collect fluorescence light from $\sim 10^4$ molecules at a time, they were able to determine the diffusion coefficient of the dye and both the binding and unbinding rates of the dye to the DNA. These were among the first direct measurements of nonequilibrium fluctuations in a system at equilibrium, and gave rise to what is today a major area of research in chemistry and biophysics. Modern implementations of FCS take advantage of the technological advances made in the last 35 years — high numerical aperture microscope objectives and single-photon sensitive detectors make it possible to study single molecules, one at a time, using FCS.

This chapter begins with an introduction to FCS, in which we derive the standard mathematical expressions for a very simple example experiment. Next we generalize these derivations to account for the systematic contributions to the FCS curve that arise when the fluorescent particles are tracked by the apparatus. Andy Berglund was the first to make such calculations[18, 37, 42]; we further generalize his ideas here in order to incorporate internal translational motion of the tracked particle into the FCS curve in preparation for our work on DNA polymer dynamics in Chapter 7. We include a discussion of the relationship between closed-loop FCS and popular approaches for studying polymer dynamics, such as dynamic light scattering or video correlation measurements. This chapter concludes with a derivation of the contribution to the FCS curve due to the three-dimensional beam modulation described in Chapter 2.

4.1 Open-loop FCS

We begin by discussing FCS in its most basic form: a stationary laser beam is used to collect fluorescence from molecules that drift into and out of it by Brownian motion. Much of this is relevant to our discussion of tracking-FCS in the next section: the notation and approach to calculating the correlation functions are the same, just much simpler in this case. We begin our discussion by establishing the notation that we will use throughout this chapter, and by introducing the correlation function in general terms.

4.1.1 The correlation function $g_2(\tau)$

We first introduce a set of functions known only as "the functions f_n" that are very useful for describing point processes[8]. In general, $f_n(t_1,\ldots,t_n)\mathrm{d}t_1\cdots\mathrm{d}t_n$ is the probability that an event (in our case, this event will always be the detection of a photon) takes place in each of the infinitesimal intervals $(t_1, t_1+\mathrm{d}t_1),\ldots,(t_n, t_n+\mathrm{d}t_n)$. In the case of fluorescence photon detections, these functions are very closely related to the measured fluorescence rate Γ^t (where the superscript t will be used extensively in this chapter to indicate a function of time). This relationship is given by

$$f_n(t_1,\ldots,t_n) = \left\langle \Gamma^{t_1}\cdots\Gamma^{t_n}\right\rangle, \tag{4.1}$$

where the average is an ensemble average that is necessary when the fluorescence rate is determined by a random process (such as the Brownian motion of a fluorescent particle). The related quantities Γ^t and f_n can be used interchangeably in calculating FCS curves, and I do so in this thesis. My choice of which to use will usually be determined by whether the dynamics I am modeling are described intuitively by a simple fluorescence rate, or whether the discrete nature of photon collection is important. In particular, in studying the statistics of photon emission in Chapter 6 I will use f_n, and I will use Γ^t otherwise. One feature of f_n that is sometimes convenient is the simple notation for expressing *conditional* probabilities: for example, $f_{1|1}(t_2|t_1)\mathrm{d}t_2$ can represent the probability of detecting a photon in time interval $(t_2, t_2+\mathrm{d}t_2)$ *given* that one was detected in $(t_1, t_1+\mathrm{d}t_1)$. A final note on the functions f_n: although they are interpreted identically to probability densities over the infinitesimal time intervals $\mathrm{d}t_n$, there is no reason for them to be normalized and so they are not probability densities themselves.

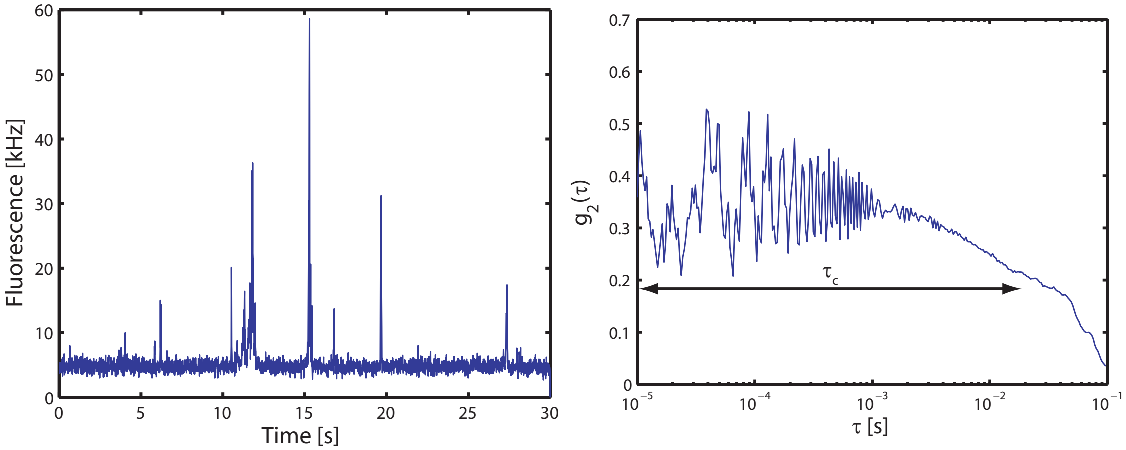


Figure 4.1: Example FCS curve (right) for open-loop FCS data (left). Freely-diffusing Qdot 655 quantum dots emit fluorescence bursts on a time-scale of about 20ms. The duration of these bursts determines the correlation time τ_c of the dominant decay of the g_2 curve. The oscillations in the FCS curve at shorter times are due to the three-dimensional beam modulation described in Sect. 2.3.

The general procedure for performing FCS is to compute the two-time correlation function of the detected fluorescence signal. In terms of f_n, the correlation function is defined as

$$g_2(\tau) = \frac{\langle f_2(t, t+\tau)\rangle}{\langle f_1(t)\rangle\langle f_1(t+\tau)\rangle} - 1, \tag{4.2}$$

where the averages are taken over t. It will usually be the case that all processes contributing to the fluorescence signal are *stationary*[8]; in this case, the time averages are not necessary. However, we leave the averages there because there are common nonstationary contributions (such as photobleaching) that we want to be able to include in our definition. It is important to note here that this function is defined with a normalization that is standard for FCS[46] but unusual in general for stochastic processes[8]. The reason for this normalization is that it eliminates systematic contributions to the correlation function at $\tau = 0$, so that $g_2(0)$ can be interpreted in a meaningful way. It is also important to note that while the FCS curve is defined as a time average, theoretical calculations often assume that the fluorescence signal is *ergodic*[8], so that time and ensemble averages are equal, and therefore use ensemble averages to compute the relevant FCS statistics. Throughout this thesis this assumption is never violated, so we will not distinguish between these two types of averages.

The correlation function expresses, in a simple form, the way in which fluctuations in a signal persist over time. Except in very unusual cases (Section 6.2.5), the correlation is greatest at $\tau = 0$; $g(0)$ is proportional to the variance of the signal. At longer values for τ, the correlation decays on time-scales that are characteristics of the fluctuations, so that $\lim_{\tau\to\infty} g(\tau) = 0$. The value at $\tau = 0$ is sometimes referred to as the *contrast* of the FCS curve because larger $g_2(0)$

will make all of the decays in $g_2(\tau)$ more pronounced. This means that parameters may be extracted from the FCS curve more accurately, so it is often desirable to design an experiment to make the contrast as large as possible.

The fluorescence bursts due to molecules drifting through the focus of the laser in a single-molecule microscopy experiment is an obvious example of the type of fluctuation that FCS is sensitive to. Each burst persists for a random amount of time, depending on the particle's Brownian trajectory, and the correlation function decays sharply on time-scales longer than the average burst duration. This is illustrated in Fig. 4.1. Features other than Brownian motion can be much less apparent in the fluorescence signal: for example, the high-frequency oscillations in the g_2 curve in Fig. 4.1 are on time-scales so short compared to the fluorescence rate that only a few photons are detected in any single period of the modulation. It is only by averaging over the entire fluorescence signal that such features can be resolved.

4.1.2 Single-component FCS

Once we have computed the correlation function in Eq. 4.2 from our fluorescence data, we must interpret that function based on a mathematical model for the dynamics of the particles that gave rise to the data. In essence, we derive the predicted $g_2(\tau)$ based on our model, and determine the values of key parameters of our model by fitting to the data. We will now derive $g_2(\tau)$ for the simplest sample possible: a single type of fluorescent particle with no dynamics other than Brownian motion. While very simple, this example illustrates the general approach to FCS and — perhaps owing to its simplicity — is one of the most common applications of the technique. For a more detailed review of FCS calculations with more general models, see [46].

We let the concentration of fluorescent particles in our sample be denoted $C(\mathbf{x}, t)$. For a sample composed of discrete particles, the concentration has the same interpretation as the rate of a Poisson process: $C(\mathbf{x}, t)\mathrm{d}^3\mathbf{x}$ is the probability that a particle can be found in an infinitesimal region of volume $\mathrm{d}^3\mathbf{x}$ about the point $\mathbf{x}$. Given this interpretation, the intensity of the fluorescence that we collect at time t is given by

$$\Gamma^t = \int \mathrm{d}^3\mathbf{x}\,\Gamma(\mathbf{x})C(\mathbf{x}, t), \tag{4.3}$$

where $\Gamma(\mathbf{x})$ is the spatial dependence of the fluorescence rate that we assume is not time-dependent, because there is no reason for us to be modulating the excitation beam in open-loop FCS. It is standard practice to now express Eq. 4.3 in terms of the Fourier transforms of $\Gamma(\mathbf{x})$ and $C(\mathbf{x}, t)$ because, although not obvious quite yet, this simplifies calculations in later steps of the derivation. In terms of the wave vector $\mathbf{k}$, Eq. 4.3 becomes

$$\Gamma^t = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\,\tilde{\Gamma}(-\mathbf{k})\tilde{C}(\mathbf{k},t), \tag{4.4}$$

where tildes denote Fourier transforms.

On our way to computing the FCS curve, we consider the quantity

$$f_2(t,t+\tau) = \left\langle \Gamma^t\Gamma^{t+\tau} \right\rangle = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}\,\tilde{\Gamma}(-\mathbf{k})\tilde{\Gamma}(-\mathbf{k}') \left\langle \tilde{C}(\mathbf{k},t+\tau)\tilde{C}(\mathbf{k}',t) \right\rangle, \tag{4.5}$$

where the averaging brackets appear over the concentration terms only, because the Γ terms are deterministic. Now we must consider the time evolution of $C(\mathbf{x},t)$. We know that $C(\mathbf{x},t)$ obeys the diffusion equation

$$\frac{\partial}{\partial t}C(\mathbf{x},t) = D\nabla^2 C(\mathbf{x},t), \tag{4.6}$$

where D is the diffusion coefficient, because we have assumed that Brownian motion alone contributes to the motion of the particles in the sample. Taking the Fourier transform of Eq. 4.6 and solving, we get

$$\tilde{C}(\mathbf{k},t+\tau) = \exp\left(-\mathbf{k}^T\mathbf{k}D\tau\right)\tilde{C}(\mathbf{k},t), \tag{4.7}$$

which we can now insert into Eq. 4.5 to get

$$f_2(t,t+\tau) = \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}\,\tilde{\Gamma}(-\mathbf{k})\tilde{\Gamma}(-\mathbf{k}') \left\langle \tilde{C}(\mathbf{k},t)\tilde{C}(\mathbf{k}',t) \right\rangle \exp\left(-\mathbf{k}^T\mathbf{k}D\tau\right). \tag{4.8}$$

Finally, with some manipulation of the Fourier transforms we get

$$\left\langle \tilde{C}(\mathbf{k},t)\tilde{C}(\mathbf{k}',t) \right\rangle = (2\pi)^6\bar{C}^2\delta^3(\mathbf{k})\delta^3(\mathbf{k}') + (2\pi)^3\bar{C}\delta^3(\mathbf{k}+\mathbf{k}'), \tag{4.9}$$

where $\bar{C}$ is the average concentration of the sample. This result follows as an application of the statistics given in Eq. 2.12 because the particles move independently of each other, and so the number of particles in any volume V is a Poisson random variable with parameter $\bar{C}V$.

We are finally in a position to compute $g_2(\tau)$, provided we insert an appropriate choice for $\Gamma(\mathbf{x})$. It is important to note that $\Gamma(\mathbf{x})$ tells us the rate at which we *detect* photons, which is not necessarily proportional to the rate at which photons are *emitted* because the FCS apparatus has an imperfect collection efficiency that depends on the particle's position in a nonlinear way. For example, the presence of a confocal pinhole in the detection optics limits the collection efficiency of the microscope along the z axis. This is generally the justification for using the three-dimensional Gaussian intensity profile

$$\Gamma(\mathbf{x}) = \Gamma_0 \exp\left[-\frac{2}{w^2}\left(x^2 + y^2 + z^2\zeta^2\right)\right], \tag{4.10}$$

where ζ accounts for the difference in the intensity decay in the axial and radial directions, in place of the Gaussian-Lorentzian profile of Eq. 2.1. In truth, the Gaussian does a great deal to simplify calculations involving $\tilde{\Gamma}(\mathbf{k}, t)$, and so I suspect its initial use was motivated by practical reasons, not by the argument presented above. In any case, this argument is true, and the Gaussian (with $\zeta \sim \lambda/\sqrt{2}\pi w$) is a good approximation to the Gaussian-Lorentzian for small z anyway. I will use the Gaussian beam profile exclusively in this chapter.

The Fourier transform of Eq. 4.10 is

$$\tilde{\Gamma}(\mathbf{k}) = \Gamma_0 \frac{\pi^3 w^3}{8\zeta} \exp\left[-\frac{w^2}{8}\left(k_x^2 + k_y^2 + \frac{k_z^2}{\zeta^2}\right)\right]. \tag{4.11}$$

Combining Eqs. 4.5-4.9 and Eq. 4.11 and inserting into Eq. 4.2, we get

$$g_2(\tau) = \frac{1}{\bar{C}\pi^{3/2}}\left(w^2 + 4D\tau\right)^{-1}\left(w^2/\zeta^2 + 4D\tau\right)^{-1/2}. \tag{4.12}$$

It is common practice to define the characteristic diffusion time $\tau_D = w^2/4D$ and the effective sampling volume $\bar{V} = \pi^{3/2}w^3/\zeta$, so that

$$g_2(\tau) = \frac{1}{\bar{C}\bar{V}}\left(1 + \tau/\tau_D\right)^{-1}\left(1 + \zeta^2\tau/\tau_D\right)^{-1/2}. \tag{4.13}$$

In this form, $g_2(\tau)$ illustrates some important features of FCS curves in general. The contrast in the FCS curve (the value $g_2(0)$) is the inverse of the effective number of particles $\bar{N} = \bar{C}\bar{V}$ that we detect simultaneously. This is quantitative justification for our claim that signals can be better resolved when fewer molecules are detected at any one time. The decay of $g_2(\tau)$ for $\tau > \tau_D$ tells us that τ_D represents the average duration of a single fluorescence burst. Figure 4.2 shows single-component open-loop FCS curves for several values of $\bar{N}$ and τ_D.

4.1.3 Incorporating additional dynamics

As we have mentioned, FCS is sensitive to much more than just Brownian motion. Any dynamics that affect the fluorescence of the particle on time-scales shorter than τ_D will in general be resolvable using FCS. Examples can include chemical reactions such as the binding and unbinding of intercalating dyes to DNA[1]; internal conformational fluctuations in polymers[21]; Förster energy transfer between donor and acceptor dyes[17]; or blinking of quantum dots[47]. Many such processes are independent of the particle's position, and affect the fluorescence

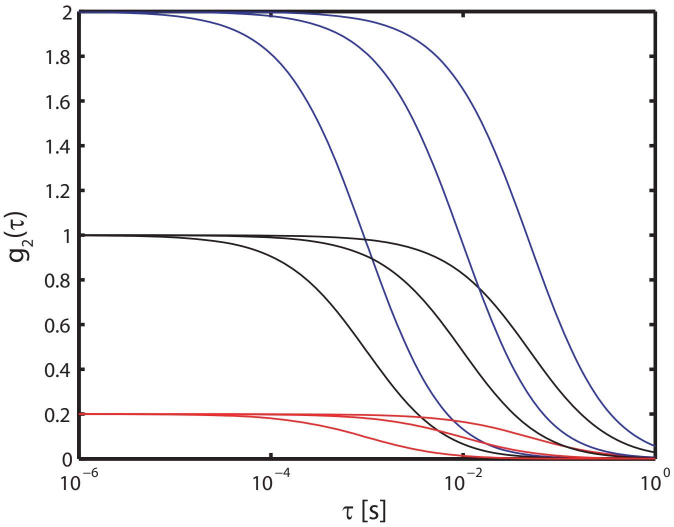


Figure 4.2: Example single-component open-loop FCS curves. Curves were plotted at effective particle numbers $\bar{N} = 0.5$ (blue), 1 (black) and 2 (red). Characteristic diffusion times were $\tau_D =$ 1ms, 5ms and 10ms. For all curves, we fixed $\zeta = 0.3$.

rate Γ^t in a multiplicative way; for example, qdot blinking can be considered a process that multiplies Γ^t by 0 when the qdot blinks off and 1 when the qdot is on. Let $\Gamma^t = \Gamma_0^t \Gamma_1^t \cdots \Gamma_n^t$, where all of the Γ_j^t terms are statistically independent. Then

$$\left\langle \Gamma^t \Gamma^{t+\tau} \right\rangle = \left\langle \Gamma_0^t \Gamma_0^{t+\tau} \right\rangle \cdots \left\langle \Gamma_n^t \Gamma_n^{t+\tau} \right\rangle \tag{4.14}$$

so that $g_2(\tau)$ factorizes:

$$g_2(\tau) = \left[g_2^0(\tau) + 1 \right] \cdots \left[g_2^n(\tau) + 1 \right] - 1. \tag{4.15}$$

This can be quite useful for simplifying calculations; in fact, even simple FCS curves are easier to calculate because Brownian motion is not correlated along orthogonal spatial axes, so the correlation function along each axis can be calculated separately. An important class of processes that cannot be factored apart in this way are those that depend on or influence the particle's position, such as the conformational dynamics of a fluorescent polymer. All such processes must be considered simultaneously with the particle's Brownian motion, as they will be when we derive the tracking-FCS curve in Section 4.2.

4.1.4 The effect of background counts

In any real experiment it is impossible to eliminate background counts from the fluorescence signal. These counts create additive contributions to the fluorescence rate, unlike the multiplicative contributions that we saw in the previous section. Here we consider how they affect the FCS curve.

We first let the background be any process with its own f_n, and we use the superscripts p to denote the fluorescence statistics of the particle and B to denote the statistics of the background. Then

$$g_2(\tau) = \frac{f_2^p(t,t+\tau) + f_2^B(t,t+\tau) - f_1^p(t)f_1^p(t+\tau) - f_1^B(t)f_1^B(t+\tau)}{f_1^p(t)f_1^p(t+\tau) + f_1^p(t)f_1^B(t+\tau) + f_1^p(t+\tau)f_1^B(t) + f_1^B(t)f_1^B(t+\tau)}, \tag{4.16}$$

where the form of the denominator reflects the fact that for any pair of detected photons, either both came from the particle, both came from background, or one photon came from each.

Equation 4.16 applies generally to any additive contribution to the fluorescence rate. In the case of the background, we will just assume that it is a Poisson process with rate Γ_B so that $f_1^B(t) = \Gamma_B$ and $f_2^B(t,t+\tau) = \Gamma_B^2$. Inserting this gives

$$g_2(\tau) = \frac{f_2^p(t,t+\tau) - f_1^p(t)f_1^p(t+\tau)}{f_1^p(t)f_1^p(t+\tau) + \Gamma_B f_1^p(t) + \Gamma_B f_1^p(t+\tau) + \Gamma_B^2}. \tag{4.17}$$

If we assume that the background is small relative to the signal from the particle at both times t and $t+\tau$, we compute the first-order Taylor expansion to get

$$g_2(\tau) \approx g_2^0(\tau)\left\{1 - \left(\frac{\Gamma_B}{f_1^p(t)} + \frac{\Gamma_B}{f_1^p(t+\tau)}\right)\right\}, \tag{4.18}$$

where $g_2^0(\tau)$ represents $g_2(\tau)$ with $\Gamma_B = 0$. This shows that the size of the FCS curve is uniformly reduced due to the background. Essentially what is happening is the FCS curve becomes a weighted average of $g_2^0(\tau)$ and the curve $g_2^B(\tau)$ for the background. For a purely Poisson background, $g_2^B(\tau) = 0$ so the effect of the averaging is to attenuate $g_2^0(\tau)$. While dynamic time-scales are preserved in $g_2(\tau)$, failure to correct for a large background will influence the interpretation of the curve at $g_2(0)$: in the single-component FCS example, it will give the appearance of a higher sample concentration. More information on FCS background correction is given in [48].

4.2 Tracking-FCS

FCS is a very useful approach to evaluating the dynamic information in fluorescence signals. This is not only true for particles probed via open-loop FCS; the fluorescence that we measure while tracking a particle can be analyzed by FCS as well, provided we are careful to incorporate the systematic differences between this scenario and open-loop FCS. In particular, both the residual motion of the particle relative to the laser due to tracking errors (as discussed in the previous chapter) and the laser modulation described in Chapter 2 contribute fluorescence fluctuations that appear in the FCS curve.

In this section we take a fairly general approach to deriving the closed-loop FCS curve. From the beginning, we incorporate internal motion of the tracked particle in a way that will allow us to calculate the FCS curves for objects ranging from large solid objects to objects with fairly complicated conformational dynamics, such as polymers. This work is a generalization of Andy Berglund's work [37, 42], in which he calculated the closed-loop FCS curve for a point particle.

4.2.1 General calculation of the tracking-FCS curve

The tracking-FCS curve is computed by following exactly the procedure in Section 4.1.2, but the quantity $\left\langle \tilde{C}(\mathbf{k}, t+\tau)\tilde{C}(\mathbf{k}', t)\right\rangle$ presents much more of a challenge than in the open-loop case. In this section we compute the function $f_2(t, t+\tau)$ in terms that are sufficiently general to apply to a variety of specific particle types, from solid objects to polymers. The calculations in this section are some of the most involved in this thesis, but I think that the simplicity with which specific results follow in later sections justify the complexity.

We allow our fluorescent particle to be labeled with N identical fluorescent emitters, so that the concentration of emitters is given by

$$C(\mathbf{x}, t) = \sum_{l=1}^{N} \delta\left(\mathbf{x} - \mathbf{x}_l^t\right), \tag{4.19}$$

where $\mathbf{x}_l^t$ is the position of emitter l at time t. Any intrinsic dynamics of the particle that we study will appear in the statistics of the $\mathbf{x}_l^t$. We let $\mathbf{x}_s^t$ be the position of the tracking stage at time t. If $\Gamma(\mathbf{x} - \mathbf{x}_s^t, t)$ is the spatially-dependent fluorescence rate, which is now time-dependent due to beam modulation, then in analogy to Eq. 4.4 the detected fluorescence rate is

$$\Gamma^t = \sum_l \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \tilde{\Gamma}(-\mathbf{k}, t) e^{i\mathbf{k}^T(\mathbf{x}_l^t - \mathbf{x}_s^t)}. \tag{4.20}$$

From this, we get the expression

$$f_2(t, t+\tau) = \sum_{l,m} \int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} \int \frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3} \tilde{\Gamma}(-\mathbf{k}, t+\tau)\tilde{\Gamma}(-\mathbf{k}', t) \left\langle e^{i\mathbf{k}^T(\mathbf{x}_m^{t+\tau} - \mathbf{x}_s^{t+\tau}) + i\mathbf{k}'^T(\mathbf{x}_l^t - \mathbf{x}_s^t)} \right\rangle. \tag{4.21}$$

In order to simplify Eq. 4.21, we make the assumption that the statistics of the terms in the exponential are all Gaussian. This is not too restrictive to prevent us from applying the result to interesting systems: it is a very good approximation for Gaussian polymers, although not exactly correct[49]. Our tracking stages are constantly moving, so there is no natural choice of coordinate axes. For convenience we may choose our coordinates so that $\left\langle x_l^t \right\rangle = \langle x_s^t \rangle = 0$

by letting the center of mass coordinate of the emitters $\mathbf{x}_{cm}^{t} = 0$. We now make use of the following property: If ξ is a constant vector and $\mathbf{x}$ is a zero-mean Gaussian vector random variable, then

$$\left\langle e^{\xi^T \mathbf{x}} \right\rangle = \exp\left[\frac{1}{2}\xi^T \left\langle \mathbf{x}\mathbf{x}^T \right\rangle \xi\right], \tag{4.22}$$

which is verified by straightforward calculation of a Gaussian integral. Applying Eq. 4.22 to Eq. 4.21, we have

$$\begin{aligned}\left\langle e^{i\mathbf{k}^T\left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)+i\mathbf{k}'^T\left(\mathbf{x}_l^{t}-\mathbf{x}_s^{t}\right)} \right\rangle &= \exp\left\{-\frac{1}{2}\mathbf{k}^T \left\langle \left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)\left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)^T \right\rangle \mathbf{k}\right. \\ &\left. -\frac{1}{2}\mathbf{k}'^T \left\langle \left(\mathbf{x}_l^{t}-\mathbf{x}_s^{t}\right)\left(\mathbf{x}_l^{t}-\mathbf{x}_s^{t}\right)^T \right\rangle \mathbf{k}' - \mathbf{k}^T \left\langle \left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)\left(\mathbf{x}_l^{t}-\mathbf{x}_s^{t}\right)^T \right\rangle \mathbf{k}'\right\}.\end{aligned} \tag{4.23}$$

Rearranging terms and reapplying Eq. 4.22 to contract the coherent terms, we have

$$\begin{aligned}&\left\langle e^{i\mathbf{k}^T\left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)+i\mathbf{k}'^T\left(\mathbf{x}_l^{t}-\mathbf{x}_s^{t}\right)} \right\rangle \\ &\quad= \left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}} \right\rangle \left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{s}} \right\rangle \exp\left\{\left\langle \left[\mathbf{k}^T\mathbf{x}_m^{t+\tau} + \mathbf{k}'^T\mathbf{x}_l^{t}\right]\left[\left(\mathbf{x}_s^{t+\tau}\right)^T\mathbf{k} + \left(\mathbf{x}_s^{t}\right)^T\mathbf{k}'\right]^T \right\rangle\right\},\end{aligned} \tag{4.24}$$

where we have defined

$$\bar{\mathbf{k}} = \begin{pmatrix} \mathbf{k}' \\ \mathbf{k} \end{pmatrix} \quad \bar{\mathbf{x}}_s = \begin{pmatrix} \mathbf{x}_s^{t+\tau} \\ \mathbf{x}_s^{t} \end{pmatrix} \quad \bar{\mathbf{x}}_{lm} = \begin{pmatrix} \mathbf{x}_l^{t} \\ \mathbf{x}_m^{t+\tau} \end{pmatrix} \tag{4.25}$$

to simplify notation.

We can greatly simplify the averages in the exponent of Eq. 4.24 by eliminating uncorrelated terms. We know that the $\mathbf{x}_s^t$ is not correlated with $\mathbf{x}_l^t$ because the fluctuations determining the positions of the two — tracking error for $\mathbf{x}_s^t$ and internal motion for $\mathbf{x}_l^t$ — are uncorrelated. Based on the same argument, $\mathbf{x}_s^{t+\tau}$ is not correlated with $\mathbf{x}_l^t$, and $\mathbf{x}_s^t$ is not correlated with $\mathbf{x}_m^{t+\tau}$. Due to our choice of coordinates, all of the correlations between these terms are zero.

We do expect there to be correlation between the remaining pair of terms. Both $\mathbf{x}_m^{t+\tau}$ and $\mathbf{x}_s^{t+\tau}$ follow the motion of the center of mass of the emitters, so their correlation (while zero for $\tau = 0$) increases steadily over time. By conditioning on and integrating over all possible values for the center of mass coordinate $\mathbf{x}_{cm}^{t+\tau}$ using the Chapman-Kolmogorov equation[8], it can be shown that

$$\left\langle \mathbf{x}_m^{t+\tau}\left(\mathbf{x}_s^{t+\tau}\right)^T \right\rangle = 2D_g\tau\,\mathrm{Id}_3, \tag{4.26}$$

where D_g is the diffusion coefficient of the center of mass coordinate and Id_3 is the 3×3 identity matrix. We will see later that the term in Eq. 4.26 will be exactly canceled by terms in the remaining two averages in Eq. 4.24, so that the center of mass diffusion coefficient disappears from the FCS curve. This represents the fact that the tracking system cancels (on average) the Brownian motion of the particle.

Inserting Eq. 4.26, 4.24 becomes

$$\left\langle e^{i\mathbf{k}^T\left(\mathbf{x}_m^{t+\tau}-\mathbf{x}_s^{t+\tau}\right)+i\mathbf{k}'^T\left(\mathbf{x}_l^t-\mathbf{x}_s^t\right)}\right\rangle = \left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}}\right\rangle \left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_s}\right\rangle e^{2\mathbf{k}^T\mathbf{k}D_g\tau}. \tag{4.27}$$

We may now simplify the average over the $\mathbf{x}_s$ terms in Eq. 4.27. To compute the average we require the joint probability density $p\left(\mathbf{x}_s^{t+\tau}, \mathbf{x}_s^t\right)$, which we can find using the Chapman-Kolmogorov equation and integrating over all possible values for $\mathbf{x}_{cm}^{t+\tau}$:

$$p\left(\mathbf{x}_s^{t+\tau}, \mathbf{x}_s^t\right) = \int \mathrm{d}^3\mathbf{x}_{cm}^{t+\tau}\, p\left(\mathbf{x}_s^{t+\tau}, \mathbf{x}_s^t | \mathbf{x}_{cm}^{t+\tau}\right) p\left(\mathbf{x}_{cm}^{t+\tau}\right). \tag{4.28}$$

In terms of the tracking error $\mathbf{e}^t = \mathbf{x}_s^t - \mathbf{x}_{cm}^t$, $p\left(\mathbf{x}_s^{t+\tau}, \mathbf{x}_s^t | \mathbf{x}_{cm}^{t+\tau}\right)$ is just the joint probability $p(\mathbf{e}^{t+\tau}, \mathbf{e}^t)$ calculated in Section 3.2.2. We insert the center of mass diffusion and integrate Eq. 4.28, and we find that the average is given by

$$\left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_s}\right\rangle = \exp\left\{-\frac{1}{2}\bar{\mathbf{k}}^T\left[\Sigma_{\mathbf{e}}^{\tau} + \begin{pmatrix} 2D_g\tau & 0 \\ 0 & 0 \end{pmatrix}\right]\bar{\mathbf{k}}\right\}, \tag{4.29}$$

where $\Sigma_{\mathbf{e}}$ is the block-diagonal matrix of tracking errors defined in Section 3.2.2. Equation 4.29 captures the contribution to the FCS curve due to tracking errors that cause the position of the center of mass of the particle to fluctuate relative to the position of the tracking stage. Such fluctuations expose the particle to variations in the excitation beam intensity, resulting in fluorescence fluctuations. The factor of $\exp\left(-\mathbf{k}^T\mathbf{k}D_g\tau\right)$ represents the fact that the stage exhibits Brownian motion because it tracks a diffusing particle, and cancels half of the term in Eq. 4.26.

I should mention an additional source of systematic tracking error here. The tracking system is only sensitive to the positions of the emitters, not to the true center of mass of the particle. Because of this, the apparatus only tracks an *estimate* of $\mathbf{x}_{cm}^t$, the variance of which scales inversely as the number of emitters in the particle. This adds a factor proportional to the variance of the distribution of particles and inversely proportional to N to the static localization figures along the diagonal of $\Sigma_{\mathbf{e}}$. This never presents a problem in the applications in this thesis, because the particles that we track are either very small compared to the tracking error or are relatively densely labeled. However, it would be an important consideration in studying, for example, large polymers labeled only on their ends.

The remaining average term in Eq. 4.27 is determined by the statistics of the particle being tracked. This will generally be of greatest interest to us because it contains the statistics that are not systematics of the apparatus, but rather dynamics of the particle we are tracking. We can simplify this term a bit if we re-expand it using Eq. 4.22 to get

$$\begin{aligned}&\left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}}\right\rangle \\ &\quad= \exp\left\{-\frac{1}{2}\mathbf{k}^T\left\langle \mathbf{x}_m^{t+\tau}\left(\mathbf{x}_m^{t+\tau}\right)^T\right\rangle\mathbf{k} - \frac{1}{2}\mathbf{k}'^T\left\langle \mathbf{x}_l^{t}\left(\mathbf{x}_l^{t}\right)^T\right\rangle\mathbf{k}' - \mathbf{k}^T\left\langle \mathbf{x}_m^{t+\tau}\left(\mathbf{x}_l^{t}\right)^T\right\rangle\mathbf{k}'\right\}. \qquad (4.30)\end{aligned}$$

Based on our assumption that the statistics of the dye positions are Gaussian, it can be shown that $\left\langle \mathbf{x}_m^{t+\tau}\,(\mathbf{x}_m^{t+\tau})^T\right\rangle = (2D\tau + R_g(m)^2)\,\mathrm{Id}_3$, where $R_g(m)$ is the standard deviation of dye m about the center of mass of the particle (sometimes called the *radius of gyration*). This means that the only term that contains any interesting *dynamic* information is the coefficient of the cross-**k** term. For now we do not make any assumptions about this term, because it will vary depending on the type of particle we are tracking, and we simply write it as

$$\varphi(l, m; t, \tau) \equiv \left\langle \mathbf{x}_m^{t+\tau}\left(\mathbf{x}_l^{t}\right)^T\right\rangle. \qquad (4.31)$$

If we finally combine Eqs. 4.21, 4.27 and 4.29 - 4.31 we get

$$\begin{aligned}\langle\Gamma^t\Gamma^{t+\tau}\rangle = \sum_{l,m}\int\frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3}\int\frac{\mathrm{d}^3\mathbf{k}'}{(2\pi)^3}\tilde{\Gamma}(-\mathbf{k}, t+\tau)\tilde{\Gamma}(-\mathbf{k}', t)& \\ \times\exp\left\{-\frac{1}{2}\bar{\mathbf{k}}^T\left[\Sigma_{\mathbf{e}}^{\tau} + \begin{pmatrix} R_g(m)^2\,\mathrm{Id}_3 & \varphi(l,m;t,\tau) \\ \varphi(l,m;t,\tau) & R_g(l)^2\,\mathrm{Id}_3\end{pmatrix}\right]\bar{\mathbf{k}}\right\}&, \qquad (4.32)\end{aligned}$$

which is the primary result of this section. We can compute $g_2(\tau)$ for a large number of interesting types of particles by inserting appropriate choices for $\varphi(l, m; \tau)$, and we can account for the systematic contribution of the beam modulation to the FCS curve (see Section 4.2.4) by inserting the appropriate $\tilde{\Gamma}(\mathbf{k}, t)$. In the event that we are using an ordinary, stationary Gaussian beam for excitation, we may insert Eq. 4.11 and evaluate the integral, leaving the result in terms of the adjustable $\varphi(l, m; t, \tau)$ only. We get

$$\begin{aligned}g_2(\tau) + 1 \propto \sum_{l,m}\prod_{\alpha\in\{x,y,z\}}&\left\{\left[\left(\sigma_\alpha^0\right)^2 + R_g(m)^2 + w_\alpha^2/4\right]\left[\left(\sigma_\alpha^0\right)^2 + R_g(l)^2 + w_\alpha^2/4\right]\right. \\ &\left. - \left[(\sigma_\alpha^\tau)^2 + \varphi(l,m;t,\tau)\right]^2\right\}^{-1/2}, \qquad (4.33)\end{aligned}$$

where the proportionality constant is found either by explicit calculation of $\langle\Gamma^{t+\tau}\rangle\,\langle\Gamma^t\rangle$ or by requiring that $g_2(\infty) = 0$.

4.2.1.1 Example: solid particle in a stationary Gaussian beam

The simplest example particle that we may study is a solid particle with identical embedded dyes that do not move relative to each other. This describes the polystyrene beads that we use as diagnostic particles for testing our apparatus in the next chapter (although these beads are spherical, we can approximate the dye distribution within them by a Gaussian without too much error so that the results of the previous section need not be revised to account for this difference). Provided the number of dyes N is large, we can ignore rotational motion of the particle because it will not be resolvable in the fluorescence signal. We will assume that the dyes are excited with a stationary Gaussian probe beam so Eq. 4.33 applies.

We let $\mathbf{r}_l$ be the vector from the particle's center of mass to dye l, so that $\mathbf{x}_l^t = \mathbf{r}_l + \mathbf{x}_{cm}^t$. Since we have assumed that the dyes have a Gaussian distribution throughout the particle and that $\mathbf{x}_{cm}^t = 0$, we have

$$\varphi(l, m; t, \tau) = \left\langle \mathbf{r}_m \left(\mathbf{r}_l\right)^T \right\rangle = R_g^2 \delta_{lm}, \tag{4.34}$$

where δ_{lm} is the Kronecker delta function and $R_g = R_g(m)$ is the standard deviation of the dye distribution and is independent of m. The fact that $\varphi(l, m; t, \tau)$ is a constant is an artifact of our choice to exclude rotational dynamics from our model; any exact expression must have $\lim_{\tau \to \infty} \varphi(l, m; \tau) = 0$. Since there are only N terms in which $\varphi(l, m; \tau)$ are nonzero (compared to $N(N-1)$ terms for which it is zero) and we have assumed N is very large, we neglect these terms in our calculation of $g_2(\tau)$. Inserting Eqs. 4.11 and 4.34 into Eq. 4.33, we get the result

$$g_2(\tau) = \prod_{\alpha \in \{x,y,z\}} \frac{\left(\sigma_\alpha^0\right)^2 + R_g^2 + w_\alpha^2/4}{\sqrt{\left[\left(\sigma_x^0\right)^2 + R_g^2 + w_\alpha^2/4\right]^2 - \left(\sigma_\alpha^\tau\right)^4}} - 1, \tag{4.35}$$

where $\left(\sigma_\alpha^\tau\right)^2$ is the time-dependent tracking error from Eq. 3.29 and we have defined $w_{x,y} = w$, $w_z = w/\zeta$. In the special case $R_g = 0$, Eq. 4.35 is exactly the result computed for point particles in [37, 42].

4.2.1.2 Example: free particles in a harmonic potential

An example of a different sort is that of N emitters that move independently and experience a force $F = -\gamma\left(\mathbf{x}_l - \mathbf{x}_{cm}\right)$ that confines them near their center of mass. This example is somewhat removed from reality, as no natural system I can think of behaves according to these dynamics. However, it is a nice illustration of the fact that we can extract the particle's internal dynamics from the correlation function.

In computing $\varphi(l, m; t, \tau)$ for this model, we again work in the coordinates $\mathbf{r}_l^t$ relative to the center of mass from Eq. 4.34, although now these depend on time. For any pair of different particles ($l \neq m$), $\varphi(l, m; t, \tau) = 0$ because the particles are all independent and they are centered about $\mathbf{r} = 0$; however, there is nonzero correlation for $l = m$. The conditional probability $p(\mathbf{r}_l^{t+\tau} | \mathbf{r}_l^t)$ is given by the well-known Ornstein-Uhlenbeck statistics[39]:

$$p(\mathbf{r}_l^{t+\tau} | \mathbf{r}_l^t) = \mathcal{N}\left[\mathbf{r}_l^{t+\tau}; \mathbf{r}_l^t e^{-\gamma\tau}, \left(1 - e^{-2\gamma\tau}\right) R_g^2 \, \mathrm{Id}_3\right] \tag{4.36}$$

which we use to directly compute the average

$$\varphi(l, l; t, \tau) = e^{-\gamma\tau} R_g^2 \, \mathrm{Id}_3 \,. \tag{4.37}$$

Again, in this example R_g does not depend on m.

We may now compute $g_2(\tau)$ quite easily, assuming a stationary Gaussian excitation beam:

$$\begin{aligned} g_2(\tau) = \frac{N-1}{N} \prod_{\alpha \in \{x,y,z\}} & \frac{(\sigma_\alpha^0)^2 + R_g^2 + w_\alpha^2/4}{\sqrt{\left[\left(\sigma_x^0\right)^2 + R_g^2 + w_\alpha^2/4\right]^2 - (\sigma_\alpha^\tau)^4}} \\ & + \frac{1}{N} \prod_{\alpha \in \{x,y,z\}} \frac{(\sigma_\alpha^0)^2 + R_g^2 + w_\alpha^2/4}{\sqrt{\left[\left(\sigma_x^0\right)^2 + R_g^2 + w_\alpha^2/4\right]^2 - \left[(\sigma_\alpha^\tau)^2 + e^{-\gamma\tau} R_g^2\right]^2}} - 1, \end{aligned} \tag{4.38}$$

in which both terms contain contributions due to systematic tracking errors, but only the second term contains information about the internal motion of the dyes within the particle. We are best able to resolve internal motion when N is small, w_α is small, and R_g is large. In the next section we will discuss more generally how these parameters influence the tracking-FCS curves.

4.2.2 General properties of tracking-FCS curves

There is an important, fundamental difference between tracking-FCS and open-loop FCS: the tracking system follows the center of mass of the emitters, meaning that any tracking errors arise from *correlated* displacements of the emitters away from the stage position $\mathbf{x}_s$. This means that tracking errors must contribute differently to $g_2(\tau)$ than uncorrelated internal motion of the particle. Specifically, as the number of emitters increases the contribution to $g_2(\tau)$ due to correlated fluctuations increases because the signal size increases, while the contribution due to uncorrelated fluctuations decreases because these fluctuations average to their equilibrium value. This difference is reflected by the trends in the contrast (the value $g_2(0)$) of the FCS curve with changes in the parameters of our models. We explore these differences here using the FCS curves we calculated in Eqs. 4.35 and 4.38. Figure 4.3 illustrates these trends, showing the effect of changes in each of these parameters on the FCS curves for both types of

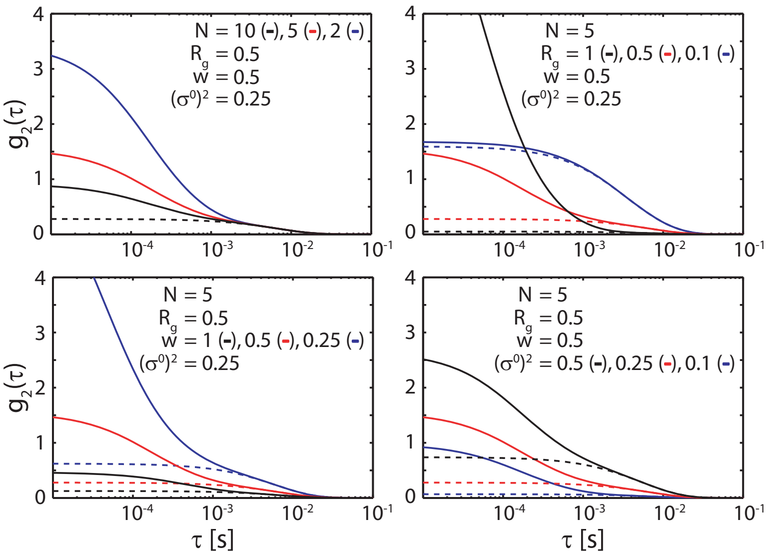


Figure 4.3: Example tracking-FCS curves for the solid particle (dashed lines) and particle containing free dyes (solid lines) with varied model parameters. In all plots, we used $\zeta = \lambda/\sqrt{2}\pi w$, $\lambda = 532nm$, $\gamma = 200$ Hz and we modeled the tracking system with a single pole at 10Hz. The remaining parameters are specified. Omitted units are all μm.

particles.

First of all, we consider how R_g affects the FCS curves. For a fixed N, increasing R_g decreases the contrast due to systematic terms. This happens because the density of dyes decreases as the particle gets larger, so synchronized tracking errors produce smaller fluorescence fluctuations. This is obvious in the case of a large solid particle, since if the particle is larger than the laser beam then no fluctuations will be visible unless the tracking error is very large. Unlike the systematic terms, the contrast due to the particle's uncorrelated internal dynamics decreases with increasing R_g because fewer emitters are detected at any one time.

We showed in Eq. 4.38 that the number of emitters N within the particle is inversely proportional to the contrast in the FCS curve due to internal motion, and this was not surprising. Whenever the motion of the individual particles is not synchronized, we expect the contrast to decrease with increasing N. However, the relative size of the contrast due to correlated tracking errors increases asymptotically in Eq. 4.38 until it reaches its value for the solid particle in Eq. 4.35. The reasoning for this is the same as before: larger N leads to larger fluctuations due to correlated motion of the emitters.

The one parameter that affects the contrast of both systematic and internal terms uniformly is the size of the laser beam. By decreasing the size of the beam, the FCS curve gains detail because fewer emitters are detected at any one time. We can generally not make the tracking laser beam very small for reasons discussed in 3.3.2, so we will often be limited in the amount of detail we can resolve in the fluorescence signal used for tracking. The need for a small beam to achieve greater contrast motivates our use of a separate probe laser beam in our measurements on DNA dynamics in Chapter 7.

Finally, in both equations the contrast due to the systematic terms scales as

$$g_2(0) \approx \frac{(\sigma^0)^4}{2\left(R_g^2 + w^2\right)^2}, \tag{4.39}$$

so that tracking errors are usually masked by the particle size and the beam size, unless those errors are very large. While the differences between correlated tracking errors and uncorrelated internal dynamics are interesting and important, it is due to this fact that we rarely see much evidence of tracking errors in the contrast of our FCS curves unless we intentionally limit tracking fidelity in order to detect them. However, in the richer FCS curves that we calculate in Section 4.2.4 in which the laser beam is modulated, tracking errors appear in other ways. For a thorough discussion the contribution of tracking errors to the FCS curve, see [42].

4.2.3 Dynamic light scattering and other related techniques

Dynamic light scattering (DLS) is a technique in which laser light is focused into a solution and objects in that solution scatter some of that light onto a detector. Fluctuations observed at the detector are used to infer properties of the objects in solution, but are usually only sufficient for measuring large-scale translational motions because of visible light's relatively small scattering amplitude off of most objects. Similarly, neutron scattering experiments illuminate target objects, sometimes containing embedded isotopic labels, and the detection statistics of the scattered neutrons are analyzed. Neutron scattering obtains higher spatial and temporal resolution than DLS, but is ineffective in aqueous samples because of the background generated by scattering off of the solvent molecules.

In both DLS and neutron scattering, the detection statistics are described by the *dynamic structure factor* of the target objects,

$$S(\mathbf{k}, t) = \frac{1}{N} \sum_{l,m} \left\langle e^{i\mathbf{k}^T \left(\mathbf{x}_m^{t+\tau} - \mathbf{x}_l^t\right)} \right\rangle, \tag{4.40}$$

where $\mathbf{x}_m^{t+\tau}$ is the location of the m^{th} scattering center. Not surprisingly, the statistics of FCS are closely related to those of scattering experiments and may be expressed in terms of $S(\mathbf{k}, t)$. Rička and Binkert showed[50] that this relationship is simply given by

$$f_2(t+\tau, t) \propto \int \mathrm{d}^3\mathbf{k} S(\mathbf{k}, \tau) \tilde{\Gamma}(-\mathbf{k}) \tilde{\Gamma}(\mathbf{k}) \tag{4.41}$$

for open-loop FCS. This is seen quite easily from our own derivations: the dynamic term

$\left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}}\right\rangle$ appearing in Section 4.2.1 is exactly equal to the average in Eq. 4.40 whenever $\mathbf{k}+\mathbf{k}' = 0$. In open-loop FCS, the delta function $\delta^3(\mathbf{k}+\mathbf{k}')$ that appears in Eq. 4.9 enforces this equality, which is precisely why Eq. 4.41 holds.

We must consider the source of this delta function term. It appears only if we assume that the sample particles are uniformly distributed throughout the sample volume, so that the probability of detecting one within any region depends only on the concentration of the sample and the size of the region. This constant probability in position-space translates into a delta-function in Fourier space.

In tracking-FCS, the probability distribution of any one of the dyes within the tracked particle does not even resemble the open-loop case; in our derivations, we have assumed that this probability is Gaussian with variance R_g^2. As a result, no delta function appears inside the integral; in fact, what does appear is the generalization of the delta function to a Gaussian with variance R_g^{-2}. It is only in the limiting case $R_g \to \infty$, corresponding either to a very large particle or to very poor tracking, that there is equivalence between $S(\mathbf{k},t)$ and $\left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}}\right\rangle$.

One application similar to tracking-FCS in which the dynamic structure factor does appear is in the use of a video camera to record the fluorescence signal. In this case, a quantity similar to the FCS curve is found by computing the autocorrelation of the fluorescence intensity in each pixel in the camera. This is the approach taken by Cohen[51], but his derivation is flawed: he requires that the point spread function of the microscope be a delta function, but as a consequence his predicted curve has a singularity at $\tau = 0$ that he does not address. We compute the correct image-image correlation function here, so that the relationship between our and his measurements is clear.

We assume a correspondence between the spatial coordinates at the camera and in solution, so that we may refer to the motion of the dye particles in the plane of the camera without ambiguity. The fluorescence intensity in a single image, as a function of the continuous two-dimensional position vector $\mathbf{x}$, is given by

$$\Gamma^t(\mathbf{x}) = \sum_l \int \mathrm{d}^2\mathbf{x}' C(\mathbf{x}',t)\Phi(\mathbf{x}-\mathbf{x}'), \tag{4.42}$$

where C is the concentration of dyes, given by Eq. 4.19, and $\Phi(\mathbf{x}-\mathbf{x}')$ is the function that maps the fluorescence of a particle centered at position $\mathbf{x}'$ in the camera frame onto the camera at position $\mathbf{x}$. This can be an arbitrary point-spread function of the microscope, except that we have not allowed for it to depend explicitly on $\mathbf{x}'$ so we rely on the assumption that the laser illumination profile is approximately constant over the region explored by the trapped particle.

We re-write this in terms of the Fourier transforms of C and Φ, and get

$$\Gamma^t(\mathbf{x}) = \sum_l \int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2} \tilde{\Phi}(\mathbf{k}) e^{-i\mathbf{k}^T(\mathbf{x}-\mathbf{x}_l^t)}. \tag{4.43}$$

If we now compute the image-image correlation function, we get

$$C(\tau) \equiv \int \mathrm{d}^2\mathbf{x} \left\langle \Gamma^{t+\tau}(\mathbf{x}) \Gamma^t(\mathbf{x}) \right\rangle = \sum_{l,m} \int \frac{\mathrm{d}^2\mathbf{k}}{(2\pi)^2} \int \frac{\mathrm{d}^2\mathbf{k}'}{(2\pi)^2} \tilde{\Phi}(\mathbf{k})\tilde{\Phi}(\mathbf{k}') \left\langle e^{i\bar{\mathbf{k}}^T \bar{\mathbf{x}}_{lm}} \right\rangle \int \mathrm{d}^2\mathbf{x} e^{-i(\mathbf{k}+\mathbf{k}')^T\mathbf{x}}. \tag{4.44}$$

The integral over $\mathbf{x}$ appears here because we compute the autocorrelation of each pixel separately rather than adding the values at all pixels and computing the autocorrelation of that quantity. The autocorrelation operation and the average-over-space operation do not commute, which is why the results end up looking different. The integral over $\mathbf{x}$ contributes a factor of $(2\pi)^2\delta^2(\mathbf{k}+\mathbf{k}')$, so that the average term — again, the same that we derived in Section 4.2.1 — collapses to the dynamic structure factor.

It is important to remember that we did not account for the variation of the excitation beam within the sample. If in fact the excitation laser did vary over the region in which the trapped particle moves, things would not have gone quite so smoothly. Equation 4.42 would contain an additional term of $\Gamma(\mathbf{x}-\mathbf{x}_s)$ in its integrand, accounting for both the illumination profile and the tracking error. This term would result in the appearance of another set of $\mathbf{k}$ vectors in Eq. 4.43; Equation 4.44 appears as it does only because the constant-intensity approximation translates into a $\delta^2(\mathbf{k}+\mathbf{k}')$ term among these additional $\mathbf{k}$ vectors. More generally, a term of the familiar form $\left\langle e^{i\bar{\mathbf{k}}^T\bar{\mathbf{x}}_{lm}} \right\rangle$ in which $\mathbf{k}+\mathbf{k}' \neq 0$ would remain in the integrand. The details of this calculation are essentially the same as in Section 4.2.1 and are not relevant to our experiments, so we need not work them out explicitly here.

4.2.4 Contributions due to modulated beams

In section 4.2.1 we computed the tracking-FCS curve in terms of an arbitrary laser excitation profile in Eq. 4.32, and in calculating two example $g_2(\tau)$ curves we assumed that the laser was a stationary Gaussian beam. Unlike the stationary beam, the modulated beams that we use for tracking make systematic contributions to $g_2(\tau)$. We derive these contributions in this section.

As we saw in Chapter 2, there are many parameters that describe the three-dimensional geometry of the two rotating tracking beams: both the beam waists and the rotation radii depend on z and can be completely different between the two beams. In computing the localization error signals, we simplified the model a little by assuming that the two beams were identical, save for an offset in z and the relative position of the waist focal plane relative to the rotation focal plane. Our calculations in that chapter were important because we needed to prove that our localization method would actually work. In this Chapter, we need not concern ourselves with as much detail in describing the beam geometry, because the tracking-FCS curves are not very sensitive to the beam geometry both because the $z-axis$ modulation is slight and because a well-tracked particle does not explore a large region of the three-dimensional laser beam.

We will approximate our excitation profile by a pair of identical three-dimensional Gaussians that are rotating at a fixed radius r and are offset in space by the distance $2z_0$. In this approximation we may say that the expected z position of the tracked particle is $z = 0$, where the beams are focused at $z = \pm z_0$. The spatially-dependent fluorescence rate relative to the stage position (in coordinates $\mathbf{e} = \mathbf{x}_{cm} - \mathbf{x}_s$) is then given by

$$\begin{aligned}\Gamma(\mathbf{e}, t) = {} & \frac{\Gamma_0}{2} \exp\left[-\frac{2}{w^2}\left(e_x - r\cos\omega_{xy}t\right)^2 - \frac{2}{w^2}\left(e_y - r\sin\omega_{xy}t\right)^2\right] \\ & \times \left\{(1-\cos\omega_z t)\exp\left[-\frac{2\zeta^2}{w^2}(e_z+z_0)^2\right] + (1+\cos\omega_z t)\exp\left[-\frac{2\zeta^2}{w^2}(e_z-z_0)^2\right]\right\}. \end{aligned} \tag{4.45}$$

In terms of the vector positions of the two laser beams

$$\mathbf{x}_1^t = \begin{pmatrix} r\cos\omega_{xy}t \\ r\sin\omega_{xy}t \\ -z_0 \end{pmatrix}, \mathbf{x}_2^t = \begin{pmatrix} r\cos\omega_{xy}t \\ r\sin\omega_{xy}t \\ z_0 \end{pmatrix} \tag{4.46}$$

and the diagonal matrix of beam waists

$$W = \begin{pmatrix} \frac{w^2}{4} & 0 & 0 \\ 0 & \frac{w^2}{4} & 0 \\ 0 & 0 & \frac{w^2}{4\zeta^2} \end{pmatrix}, \tag{4.47}$$

we may write the Fourier transform of the fluorescence profile

$$\begin{aligned}\tilde{\Gamma}(\mathbf{k}, t) = {} & \frac{\Gamma_0 \pi^3 w^3}{8\zeta}\left[\left(1-\cos\omega_{xy}t\right)\exp\left(i\mathbf{k}^T\mathbf{x}_1^t - \frac{1}{2}\mathbf{k}^T W\mathbf{k}\right)\right. \\ & \left. + (1+\cos\omega_z t)\exp\left(i\mathbf{k}^T\mathbf{x}_2^t - \frac{1}{2}\mathbf{k}^T W\mathbf{k}\right)\right]. \end{aligned} \tag{4.48}$$

In order to compute $g_2(\tau)$, we now just need to insert Eq. 4.48 into Eq. 4.32 and work through a very long sequence of algebra to simplify the result. I will spare these messy details; the

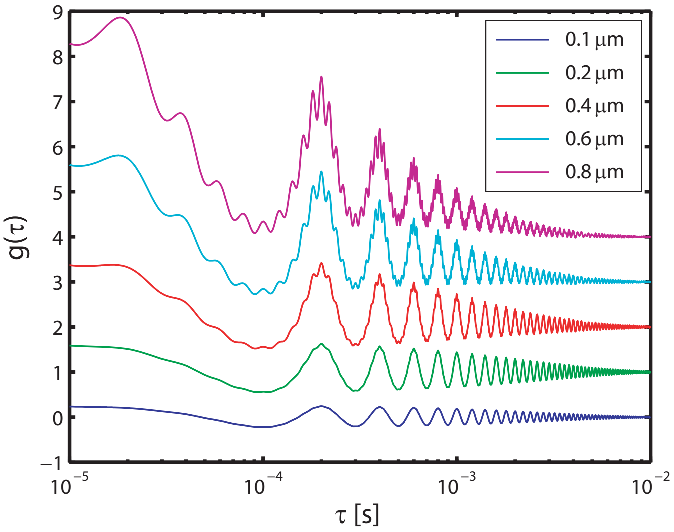


Figure 4.4: Tracking-FCS curves for a point particle with 3-D modulated laser beams and varied tracking errors. The curves are successively offset by 1 for clarity. We set $\sigma_x^0 = \sigma_y^0 = \sigma_z^0$ to the values given in the legend. Other parameters were fixed at $R_g = 0$, $w = 1\mu$m, $\zeta = \lambda/\sqrt{2}\pi w$, $\lambda = 532$nm, $r = w/\sqrt{2}$, $z_0 = 1\mu$m. We modeled the tracking system with a single pole at 50Hz. The modulation frequencies were $\omega_{xy} = 5$kHz and $\omega_z = 50$kHz.

result for the solid particle is

$$g(\tau) = \frac{\bar{\sigma}_0^4 \bar{s}_0^2}{2(\bar{\sigma}_0^4 - \sigma_\tau^4)\sqrt{\bar{s}_0^4 - (\sigma_z^\tau)^4}} \exp\left[-r^2\left(\frac{\bar{\sigma}_0^2 - (\sigma_x^\tau)^2 \cos\omega_{xy}\tau}{\bar{\sigma}_0^4 - (\sigma_x^\tau)^4}\right) + \frac{r^2}{\bar{\sigma}_0^2} + \frac{z_0^2}{\bar{s}_0^2}\right]$$
$$\times\left[\left(1 + \frac{\cos\omega_z\tau}{2}\right)\exp\left(-\frac{z_0^2}{\bar{s}_0^2 + (\sigma_z^\tau)^2}\right) + \left(1 - \frac{\cos\omega_z\tau}{2}\right)\exp\left(-\frac{z_0^2}{\bar{s}_0^2 - (\sigma_z^\tau)^2}\right)\right] - 1, \quad (4.49)$$

where we have assumed $(\sigma_x^\tau)^2 \approx \left(\sigma_y^\tau\right)^2$ and we have defined

$$\bar{\sigma}_\tau^2 = (\sigma_x^\tau)^2 + R_g^2 + \frac{w^2}{4} \quad (4.50)$$
$$\bar{s}_\tau^2 = (\sigma_z^\tau)^2 + R_g^2 + \frac{w^2}{4\zeta^2}. \quad (4.51)$$

Figure 4.4 shows $g_2(\tau)$ from Eq. 4.49 with variations in the RMS tracking error. The size of the oscillations at both modulation frequencies increases with increasing tracking error because, as shown in Chapter 2, it is proportional to the particle's distance from the origin. The oscillations decay in amplitude at longer times because of the combination of dephasing due to the particle's motion in the rotating beam and tracking error correction by the feedback system.

The term in Eq. 4.49 due to the beam rotation is identical to that in two-dimensional tracking as shown in [42]. That reference also shows how the localization of the particle along the x

and y axes is inferred from the size of the oscillations. Here we apply the same approach to the z axis. We evaluate the quantity $[g_2(0)+1]/[g_2(\pi/\omega_z)+1]$, which gives the ratio of the size of the oscillation at its maximum value to that at its minimum value. We use $\sigma_\alpha^{\pi/\omega_z} \approx \sigma_\alpha^0$ for all axes α because the modulation is much faster than the particle's motion. If we assume that $\omega_{xy} \ll \omega_z$ so that oscillations at ω_{xy} only begin to appear at times much longer than π/ω_z, we have

$$\frac{g(0)+1}{g_2\left(\pi/\omega_z\right)+1} = \frac{3\exp\left(-\frac{z_0^2}{2\left(\sigma_z^0\right)^2+R_g^2+w^2/4\zeta^2}\right)+\exp\left(-\frac{z_0^2}{R_g^2+w^2/4\zeta^2}\right)}{\exp\left(-\frac{z_0^2}{2\left(\sigma_z^0\right)^2+R_g^2+w^2/4\zeta^2}\right)+3\exp\left(-\frac{z_0^2}{R_g^2+w^2/4\zeta^2}\right)} \tag{4.52}$$

$$= 1 + \frac{z_0^2}{\left(R_g^2+w^2/4\zeta^2\right)^2}\left(\sigma_z^0\right)^2 + O\left[\left(\sigma_z^0\right)^4\right]. \tag{4.53}$$

If we isolate the oscillations at ω_{xy} by setting $\omega_z \ll \omega_{xy}$, we find that they scale the same way:

$$\frac{g(0)+1}{g_2\left(\pi/\omega_{xy}\right)+1} = 1 + \frac{2r^2}{\left(R_g^2+w^2/4\right)^2}\left(\sigma_x^0\right)^2 + O\left[\left(\sigma_x^0\right)^4\right]. \tag{4.54}$$

From Eqs. 4.53 and 4.54 it is evident why the oscillations at ω_z are so much smaller than those at ω_{xy} in the plot in Fig. 4.4. Since $w/\zeta \sim \lambda/\sqrt{2}\pi w$, with waist sizes near $w \approx 1\mu$m we expect the oscillations at ω_z to be as much as 100 times smaller than those at ω_{xy} for a given mean-squared tracking error. While it is possible to infer the tracking error by solving Eq. 4.53 for σ_z^0, the oscillations in $g_2(\tau)$ are often so small that they cannot be used for this purpose.

Chapter 5

Experimental apparatus and diagnostic results

In this chapter, we give a detailed description of the experimental apparatus that we designed to implement the three-dimensional localization technique described in Chapter 2 and to subsequently track fluorescent particles using feedback as described in Chapter 3.

5.1 Design overview

We begin with a general overview of the design of the apparatus. This section (I hope) contains enough basic information to describe the complete functioning of the tracking apparatus without requiring a background in optics. Later sections contain more detailed information.

5.1.1 Modulation optics

The most important part of our experimental apparatus is the set of optics used to modulate the laser beams for three-dimensional localization as discussed in Section 2.3 and illustrated in Fig. 2.4. Two separate laser beams must be rotated in circular orbits while the optical power in the beams is modulated with opposite phase. These beams are then combined and focused together into a sample, with the two beams focusing at slightly different depths. In this section, we discuss how all of this is accomplished.

We rotate the beams and modulate the optical power with acousto-optic modulators (AOM). These are standard optical components, but we describe their basic operation here due to their relative obscurity in other fields. For more information, see [52]. The type of AOM that we use is illustrated in Fig. 5.1. It consists of a slab of glass with a piezoelectric actuator — a crystal that expands when a voltage is applied to it — bonded to it on one side. A high-power sinusoidal RF signal (in our case $f \sim 40$MHz and $P \sim 1$W, although the power is enhanced by a resonance in the AOM's input coupler) is applied to the piezo, which creates a compression wave that

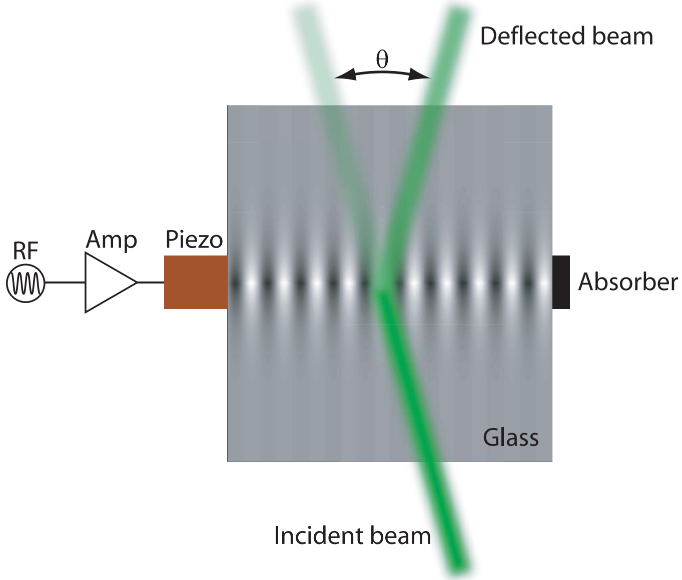


Figure 5.1: Acousto-optic modulator. High-power RF applied to a piezoelectric transducer sets up a traveling acoustic wave in a glass slab. Incident light scatters off of the resulting periodic refractive index at an angle θ depending on the RF frequency and optical wavelength.

propagates through the glass. An acoustic absorber is bonded on the opposite side of the glass to prevent reflection of the wave (actually, in our case the glass is cut at an angle to achieve this purpose). Inside the glass, the compression wave induces a sinusoidally-varying index of refraction due to the dependence of the index on the density of the glass. When incident light interacts with this index variation, some fraction of the beam is scattered at angle θ from the incident beam. This angle depends approximately linearly on the frequency of the acoustic wave, and the optical power in the deflected beam depends linearly on the acoustic wave power. Careful alignment of the laser beam to the AOM is necessary in order to couple as much of the incident beam as possible into the deflected beam — the optimal coupling angle depends on both the wavelength of the beam and the frequency of the acoustic wave. With a well-aligned 532nm beam, we have been able to couple as much as 85% of the incident optical power into the deflected beam.

We deflect our beams by varying the frequency of the RF signal applied to the AOMs. We use voltage-controlled oscillators (VCO) to perform this frequency modulation. These devices output RF at a frequency proportional to their input voltage; we drive the VCOs with a sine wave, producing the frequency-modulated RF that causes the AOM to deflect the laser beam back and forth. Each AOM produces deflection along a single Cartesian axis, so we must use two AOMs along orthogonal axes — driven with the sine and cosine quadratures of the rotation frequency — for complete beam rotation.

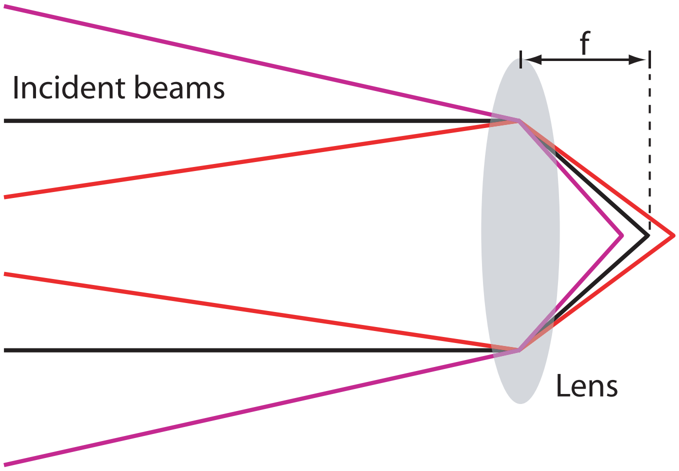


Figure 5.2: Focusing of laser beams. Colored lines represent the outer boundaries of three laser beams. Collimated beams (black) focus at almost exactly the focal distance f from the lens. Diverging beams (red) focus further from the lens, and converging beams (violet) focus nearer to the lens.

Three-dimensional localization requires two rotating beams focused at different planes in the sample. We create two beams from a single laser by linearly polarizing its output at an adjustable angle using a birefringent quartz plate known as a half-wave or $\lambda/2$ plate. A polarizing beamsplitter (PBS) separates this beam into two beams with orthogonal polarizations, equivalent to taking orthogonal projections of the incident beam's polarization onto the coordinate axes defined by the PBS. Once split, we modulate the optical power in each beam between 0 and 100% using an AOM that we drive with a sinusoidally-attenuated RF signal. Our attenuation electronics ensure that the power modulation is done with opposite phase between the two beams in order to ensure high-quality z-axis localization signals.

In addition to modulation with opposite phases, the two beams must be focused at different depths in the sample. Fig. 5.2 illustrates how this is accomplished. Converging beams focus nearer to the focusing lens than diverging beams, so we may separate the two tracking laser beams simply by mis-matching their divergence angles. This can be done by passing each beam through a pair of lenses and adjusting the spacing between either lens pair. Once these adjustments are made, the beams are recombined on a polarizing beamsplitter and are ready to focus into the sample.

Figure 5.3 contains a photo of the apparatus, showing the beam paths and all of the modulation optics. A single AOM deflects the output of the laser along the y axis. This deflected beam is split and two AOMs, one for each beam, perform the x-axis deflection and power modulation simultaneously. Pairs of lenses adjust the divergence of each beam, and the beams are recombined on a polarizing beamsplitter. Once combined, the beams are transmitted to the sample by the optics described in the next section.

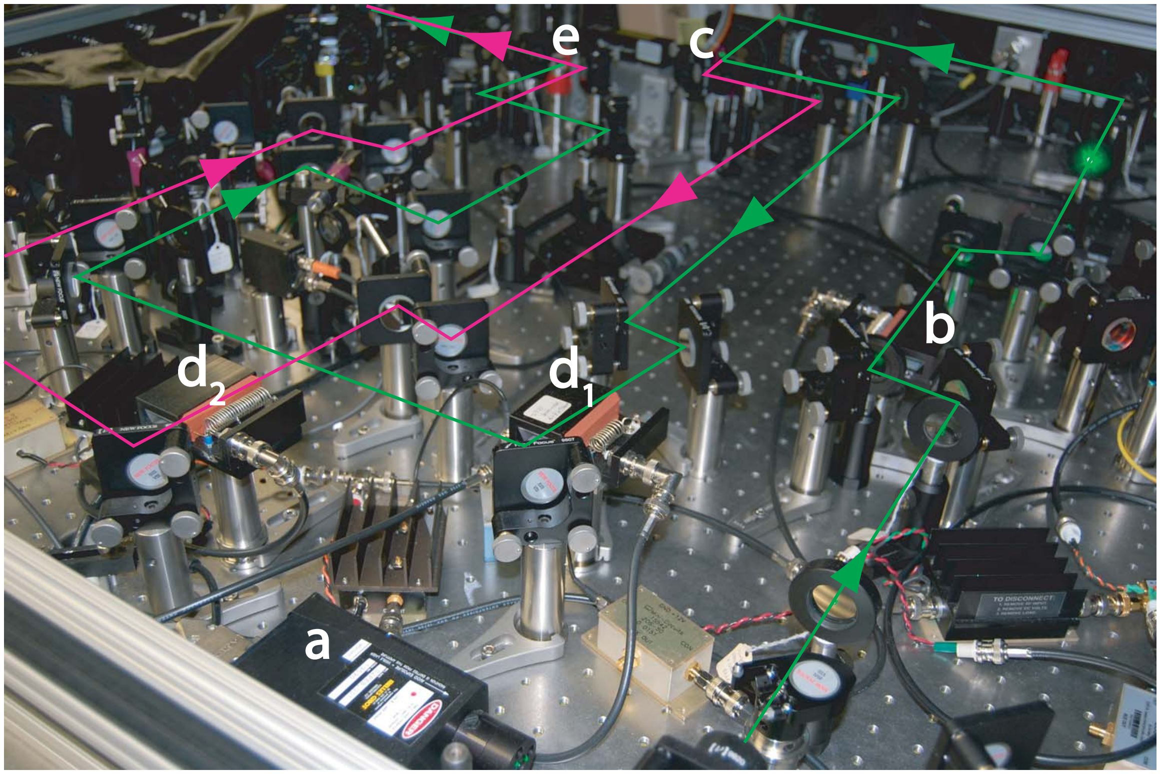


Figure 5.3: Photo of modulation optics with beam paths drawn in. Different colors distinguish the paths of the two identically-colored tracking beams. The beam originates at the laser at point **a**. It is deflected parallel to the optical table surface by the acousto-optic modulator at point **b**, then rotated 90° by a set of mirrors (not labeled) to produce the y-axis deflection. The beam is split in two at point **c**. Acousto-optic modulators at points $\mathbf{d}_1$ and $\mathbf{d}_2$ deflect along the x axis and modulate the optical power in the two beams. The beams are recombined by a polarizing beamsplitter at point **e** after passing through focusing optics (not labeled). The 1in spacing of the holes on the optical table serves as a size scale reference.

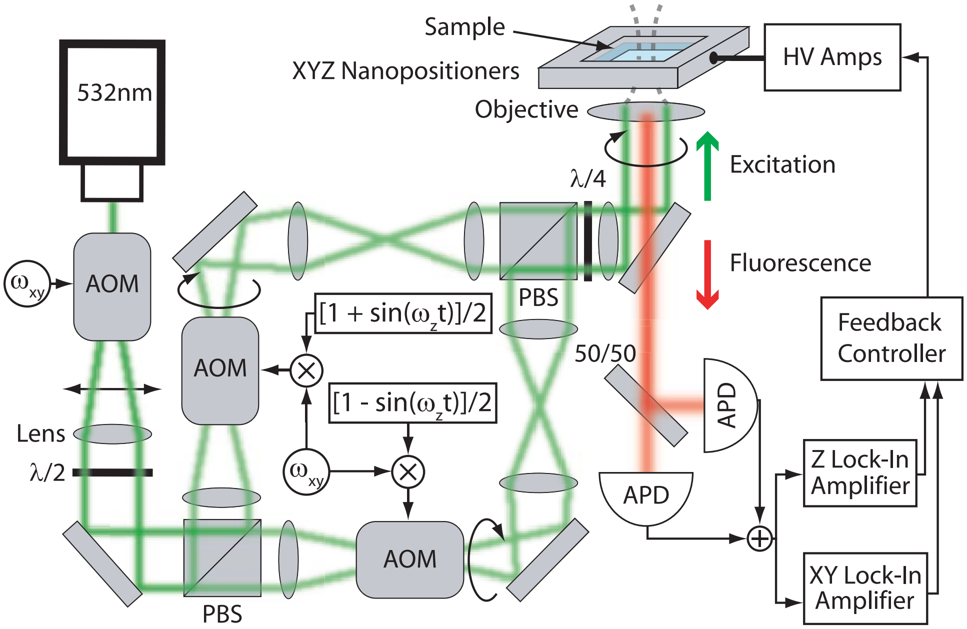


Figure 5.4: Schematic of the apparatus. Circled ω_{xy} indicates the RF output of a VCO (with 40MHz center frequency) that is frequency-modulated at ω_{xy} by driving the control port of the VCO with a sine wave. Arrows overlaid on beams in the schematic illustrate beam deflection: the beam is deflected along the vertical (y) axis by the first AOM and a set of mirrors (not shown). The remaining AOMs deflect on the horizontal (x) axis to complete the beam rotation.

5.1.2 Additional optics, detectors and feedback electronics

In the previous section, we described all of the optics necessary for modulating the tracking laser beams. We now complete our overview of the apparatus by describing the optics that transmit the beams into the sample, collect the fluorescence light, derive localization signals and actuate the tracking stages.

Figure 5.4 shows a nearly complete schematic of the apparatus, including the modulation optics described in the previous section. Once the two tracking beams are recombined, a quarter-wave (λ/4) plate is used to circularize the polarization of the two beams simultaneously. This serves primarily to ensure that each focused beam is symmetric, but additionally causes the back-reflection of either beam to lie along the beam path of the opposite beam. This aids in aligning the microscope. The two combined beams are reflected off of a dichroic mirror and focused into the sample by a microscope objective.

Fluorescence light is collected by the same microscope objective that focuses the beams and is separated from the laser light by the dichroic mirror. As shown in Fig. 5.4, the fluorescence light is split by a 50/50 beamsplitter and collected by a pair of photon-counting avalanche photodiodes (APD). This arrangement is widely known as the Hanbury-Brown-Twiss (HBT) configuration, named for the scientists who invented it for studying fast correlations in coherent light in the 1950s[53, 54]. After detecting a photon, electron cascade detectors such as photomultiplier tubes and photon-counting APDs become insensitive to subsequent photons for a brief period of time — on our APDs, about 50ns. However, a pair of detectors looking at the same fluorescence signal is sensitive to photon pairs arriving with infinitesimally short time-spacing: there is a 50% chance that the second photon will be incident on a different detector than the first, so that the first detector's dead time does not prevent its detection. We use the HBT configuration for our fast time-scale measurements in Chapter 6, but the time resolution it provides is not necessary in other measurements and so we often use only a single detector.

The fluorescence signals output by the APDs are combined and fed directly into the lock-in amplifiers. These devices input an analog signal, digitize it, and perform precisely the computation discussed in Section 2.2.2 in real-time. The result is converted back into an analog signal and fed into a feedback controller. The controller consists of three separate, independent analog integrators — one for each Cartesian axis. The outputs of the integrating controllers go directly to high-voltage amplifiers that drive the piezoelectric nanopositioner stages. These stages move the microscope objective (along the x and y axes) and the sample (along the z axis) in order to track the motion of the fluorescent particles we detect.

5.1.3 Probe beam

The apparatus we have so far described in this section is capable of tracking diffusing fluorescent particles while making high time-resolution measurements of the fluorescence signal. For many applications — for example, measuring photon statistics on individual fluorescent emitters as we do in Chapter 6 — no additional components are required. However, as we discussed in Section 4.2.2, the tracking beam is often insufficient for providing high spatial resolution for measurements on particles with internal translational degrees of freedom, such as polymers. The sensitivity of such measurements depends on the size of the focused beam inside the sample — the smaller the beam, the better the spatial resolution. As discussed in Section 3.3, tracking places exactly opposite demands on the beam size: smaller beams improve absolute tracking fidelity but impose limits on the size of the particles that can tracked. To get around this complication, we use a second, tightly-focused laser beam for measurements on internal motion while tracking with a much larger beam.

5.2 Detailed component descriptions

In the previous section we described the basic functionality of the three-dimensional tracking apparatus with a level of detail sufficient for understanding the experimental measurements appearing later in this thesis. In this section, we provide much more detail — the types of details that someone building or using this type of apparatus might find useful but few others will. This section is ordered primarily as a discussion of our experiences with the particular components used in the tracking apparatus. Many of these components were inherited from earlier generations of the apparatus, built by Andy Berglund. Descriptions of those components can be found in Andy's thesis[37]. Nonetheless, I have more to say about most of them, and I include my thoughts here.

5.2.1 Lasers

Green Tracking laser. The Melles-Griot DPSS 532nm laser is not stabilized for intensity or polarization, but I have found that the beam power delivered to the back of the microscope objective fluctuates less than 5% over time-scales of several hours. The laser did need time to warm up in order to achieve such stability: fluctuations are visible by eye for about 10 seconds after power-up, but steady-state power output seems to be reached in about 1 hour. This laser began operating intermittently at the end of the data-taking that resulted in Chapter 7 of this thesis (after six years of service), and will likely be replaced before proceeding further with more experiments.

Violet probe laser. The violet laser system is illustrated in Fig. 5.5. Despite its complexity, in most regards this setup is actually fairly robust. The Verdi V8 and MBR-110 require little maintenance — nothing more than occasional coupling realignment. The MBD-200 is also fairly stable, holding its cavity lock for many hours at a time, but its input coupling needed optimization every few days. Its output suffers from intensity fluctuations of a few percent on $\sim$ 100ms time-scales, and etalon effects in the PM fiber cause large ($\sim$ 10%) fluctuations on similar time-scales. Stability of the probe power is essential, as any amplitude noise is visible in the fluorescence correlation function. To eliminate all of this noise, I used a moderately fast ($\sim$ 3kHz -3dB bandwidth) feedback circuit comprised of an IntraAction AOM for power modulation, a New Focus amplified photodiode on the fiber output, and an analog integrating controller with a stabilized voltage reference (LM399) for setting the intensity lock point.

This configuration was chosen because it was the fastest way to generate blue/violet light since we already owned all of the necessary components. Its complexity is a major draw-

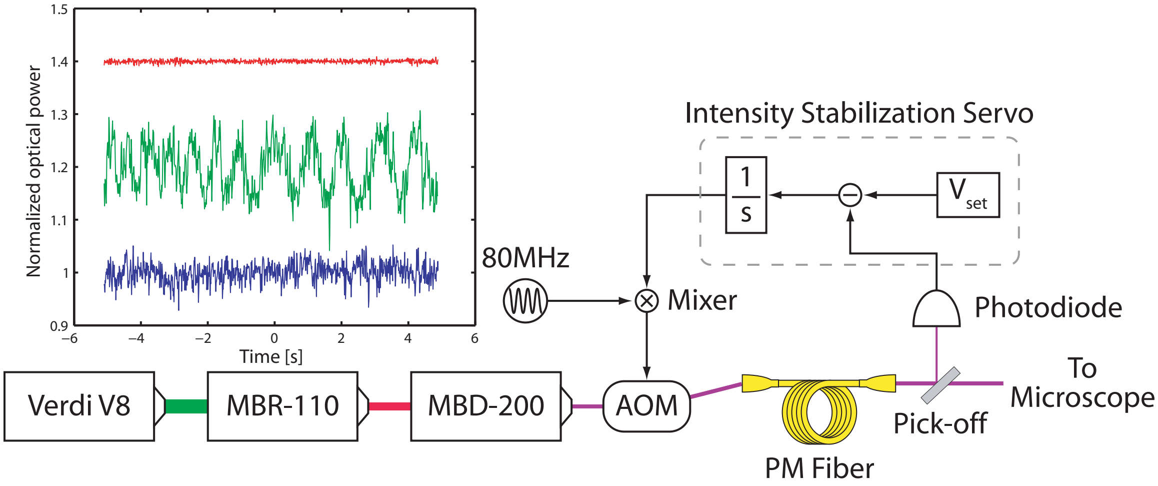


Figure 5.5: 415nm probe laser system. A Coherent Verdi V8 DPSS laser produces 5.25W at 532nm and pumps an MBR-110 Ti:S laser that produces 400mW of light at 830nm. The 830nm laser pumps an MBD-200 resonant frequency doubler that produces about 5mW of 415nm light that is coupled to a Thorlabs/Nufern polarization-maintaining fiber. The fiber's output power is stabilized using a feedback loop consisting of a New Focus photodetector, an IntraAction AOM and a home-built integrating feedback controller. Inset shows plots of typical normalized output power of the doubler (blue), fiber without feedback (green) and fiber with feedback (red), offset by 0.2 from each other for clarity.

back as it requires a relatively large amount of maintenance, a fiber to clean up its beam shape (and guide it through the lab to the tracking apparatus), and control electronics to keep its output power stable. Furthermore, it occupies almost as much optical table space as, consumes as much energy as, and was almost certainly more expensive than the entire tracking apparatus. Compact, efficient, inexpensive solid-state lasers are now available in many blue wavelengths and should always be chosen over this configuration unless tunability to very specific optical frequencies is needed (which it never is for room-temperature fluorescence spectroscopy in liquids).

5.2.2 Modulators

Acousto-Optics. The acousto-optic modulators (IntraAction AOM-40) have worked very well for beam rotation and power modulation. They are much more reliable and easily adjustable than the resonant galvanometer mirrors used for beam rotation in the two-dimensional tracking apparatus. They are also much more pleasant to work around, considering the large difference in the response of the human ear at their 40MHz (vs. 8kHz) drive frequency.

The speed of the acoustic wave inside the AOM determines the deflection fidelity: the wave must cross the entire diameter of the beam at an approximately constant frequency in order to avoid producing a blurred deflected beam. In our glass AOMs the speed of sound is approximately 3km/s, so the time needed for the acoustic wave to fully cross our $\sim$ 1mm beams is about 300ns. This means that frequencies as high as a few hundred kHz should produce high quality, nearly blurr-free deflection. In practice, we have used modulation frequencies ω_{xy} up to 75kHz without any noticeable degradation in beam quality.

During alignment, the laser beams are passed through the AOMs as close to the piezo transducers as possible without clipping — usually about 1mm. This does not affect coupling efficiency but reduces phase lag between the input RF and output beam power/deflection dramatically. With 40MHz glass AOMs, the 90° phase-lag frequency can be pushed to well-past 100kHz, allowing for the possibility of very fast feedback on the optical power or beam deflection.

We must be careful when coupling the deflected output of one AOM into another. As discussed in Sect. 2.3, the ideal beam rotation has a conical shape in three dimensions, which requires that the deflections along the two axes focus in approximately the same plane. We use lenses to image the deflection of the first AOM into the plane of the acoustic wave of the second AOM in order to achieve this goal.

One drawback to using the AOMs to rotate the beam is that the optimal alignment angle for coupling into the first-order deflected beam changes as a function of acoustic wave frequency. This results in the creation of amplitude noise on the tracking beam with a spectral peak at twice the rotation frequency, which is generally too fast to eliminate using feedback. Fortunately, we usually only need very small deflections so this effect is of little consequence. However, if we were to start feeding back to the AOMs for tracking fast particles or using rotation-radius modulation, this issue should be addressed. I have considered a few ideas for feed-forward noise cancellation, but I suspect the best solution is to buy specialized acousto-optic *deflectors*. These correct for acoustic frequency changes by (I suspect) either using multiply-resonant input coupling networks with higher Q for off-center RF frequencies or by using multiple piezos that actuate along slightly different axes and have spectrally-separated frequency responses.

RF amplitude modulators. Most devices sold as RF amplitude modulators do not respond at control frequencies higher than a few tens of kHz. I spoke to an RF engineer from the company Tele-Tech about this, and he related some useful information to me. Standard current-controlled RF attenuators are essentially just mixers with poor high-frequency response on their control ports that reduces the creation of harmonics of the input RF signal on the attenuated output. If higher-frequency harmonics are not an issue — for example, resonant AOMs respond very weakly to them — then a mixer may be substituted for the attenuator.

I was told: “a mixer is a symmetric device!” Application of AC signals to any two ports of a mixer results in the output of the product of those signals on the third port. The only functional differences between ports, in fact, are the frequency responses: often the RF and LO ports do not respond to DC, but the IF port almost always does. By driving the IF port of a mixer with a DC-offset 100kHz sine wave, we create amplitude modulation of the input RF. Driving the AOMs with this signal causes the amplitude modulation we require for z-axis localization.

In principle, a mixer can be used in this configuration for power modulation at frequencies all the way down to DC. However, when using a mixer for low-frequency beam power control I discovered an interesting problem related to the frequency modulation of the RF signals used for beam deflection. A portion of the frequency-modulated RF is visible to the electronics used to drive the mixer’s IF port. By interfering with those electronics, this can cause substantial amplitude modulation on the mixer’s output. This effect was small, between 5 and 10% of the DC power level, when the mixer was driven with an Agilent DC power supply, and twice as large when driven by an op amp. I found no improvement by replacing the mixer with a true current-controlled RF attenuator, but fortunately a solid-state voltage-controlled attenuator (Mini-Circuits ZX73-2500) provided sufficient isolation between the RF and DC signals to eliminate the effect. It is possible that sufficient isolation could also have been achieved using a passive filtering network between the electronics and the mixer, but I did not attempt to do this.

Electro-optic phase modulator. My original design for the optical power modulation used for z-axis localization used an electro-optic modulator (EOM) made by New Focus, driven on a resonance induced using an inductor in series with the EOM’s capacitance. The main reason for choosing this approach was that I was concerned about optical power losses, and it conserved power by alternating all of it between orthogonal polarizations. The EOM required temperature stabilization because its birefringence exhibited large thermal drifts, so I glued a thermistor directly to the crystal inside of it and placed a thermoelectric cooler (TEC) on the outside, as close to the crystal as possible. Driving the TEC in closed-loop to stabilize the thermistor’s resistance, the drift in the EOM’s output was greatly suppressed. I recorded my first 3-D tracking data with the EOM modulation. However, the EOM needed adjustments every week or so, and removing it in favor of doing all modulation with the AOMs greatly improved the robustness of the apparatus (at the expense of a 50% loss in optical power, which I had deemed insignificant by this point).

5.2.3 Optics

Microscope objective and collection optics. We use a Carl Zeiss C-Apochromat water-immersion (NA 1.2) microscope objective in inverted epifluorescence configuration. This objective was chosen because it is optimized for deep aqueous samples; in comparison to our Zeiss Plan Apochromat oil-immersion objective, both objectives produced good localization signals on fluorescent beads immobilized on glass surfaces but we were unable to track particles in 3-D with the oil-immersion objective. We attribute this to the aberrations induced by the refraction index mismatch between the immersion oil and the water. The C-Apochromat is very heavy (240g, compared with 140g for the Plan Apochromat) due to its many optical elements. This added weight reduces the resonance frequency of the piezo nanopositioner that moves the objective, and so effectively limits the bandwidth with which we can feed back. This has not been a problem for tracking particles as small as 20nm quantum dots in water, and can be overcome in the future by one of the strategies described in Section 3.3.2.

The red fluorescence light (from the green excitation laser) collected by the microscope objective is focused by a Zeiss $f = 160$mm tube lens, then collimated by an Edmund Optics achromat and focused again by a second achromat onto the APD active area. This configuration was unnecessary for the experiments described in this thesis, but the highly-corrected tube lens will enable the use of a confocal pinhole for reducing background fluorescence in future experiments. In contrast, the only imaging optic for the blue fluorescence light (from the violet excitation laser) is a single Edmund Optics achromat. In the future this will be replaced by a second tube lens, for integration of a confocal pinhole.

Polarization optics. For convenience and to conserve optical power, we split our beam by rotating its polarization with a zero-order half-wave plate (CVI Laser) and separating orthogonal polarizations. The power balance between these beams is critical because even small changes affect the z-axis localization signals. Fortunately, we see no evidence of drift with the CVI waveplate and New Focus polarizing beamsplitters. Once the beams are recombined, we use a multi-order quarter-wave plate (VLOC) to create circularly-polarized beams (with opposite helicity). This serves the purposes both of enforcing radial symmetry of the focused beams and of transmitting retro-reflected light from one beam along the beam path of the other, aiding beam alignment.

Filters.

We use interference filters for combining laser beams and separating fluorescence based on wavelength. These filters are comprised of a set of bandpass emitter filters from

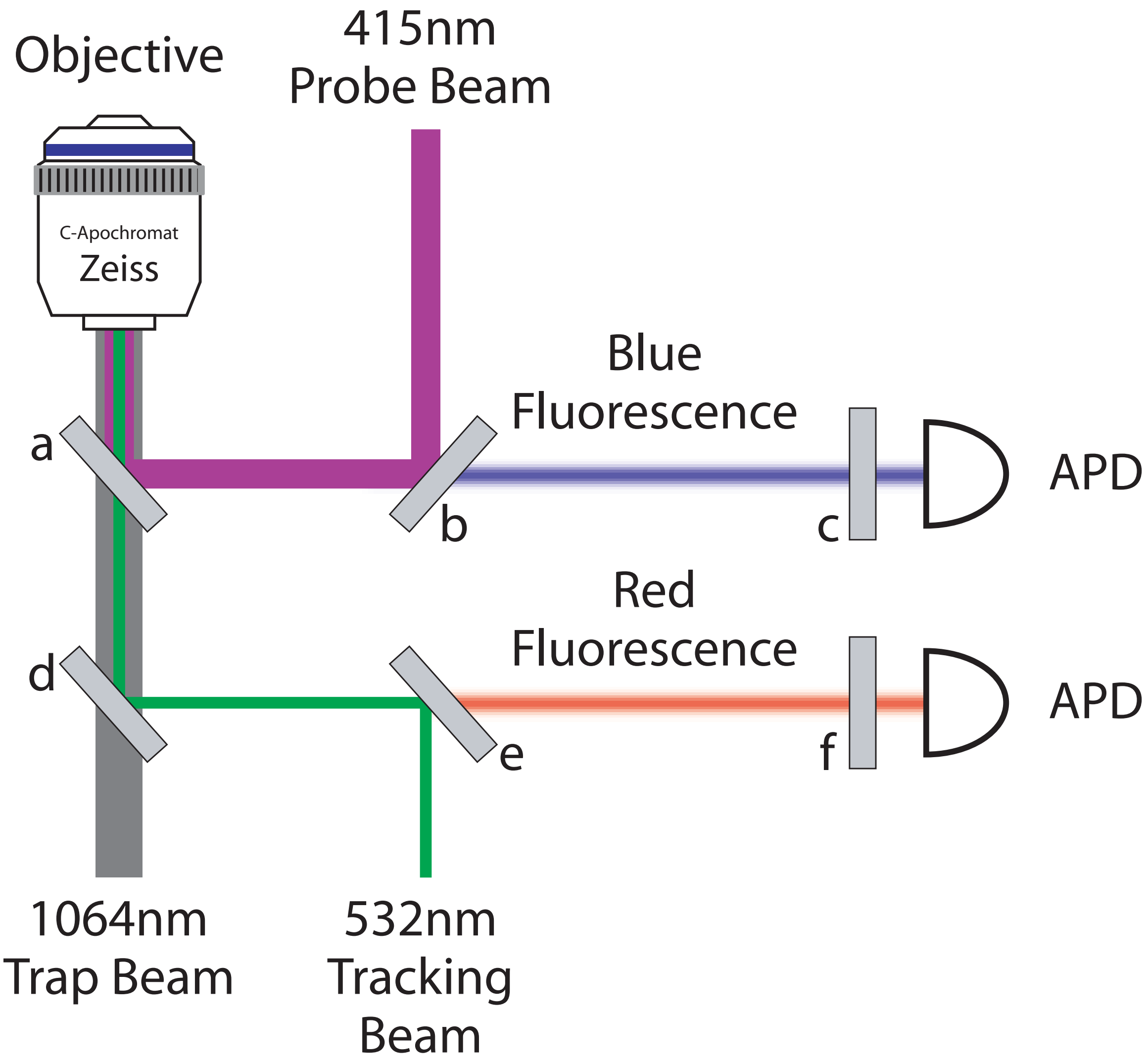


Figure 5.6: Layout of filters for light combining and separating. **a**: Omega Optical XF2031, reflects $\lambda < 505$nm. **b**: Omega Optical XF2006, reflects $\lambda < 445$nm. **c**: Chroma Technologies ET480/40m, transmits $460\text{nm} < \lambda < 500\text{nm}$. **d**: CVI Laser BSR51-1037, reflects $500\text{nm} < \lambda < 600\text{nm}$. **e**: Chroma Technologies DRLP545, reflects $\lambda < 545$nm. **f**: Chroma Technologies HQ575/50m, transmits $550\text{nm} < \lambda < 600\text{nm}$. We sometimes use a Chroma Technologies HQ630/140m-2p bandpass emitter filter on the red fluorescence channel (not shown) for its strong rejection of light at both 415nm. Note that the 1064nm beam is not used in experiments described in this thesis, but dictated the use of the CVI bandpass filter.

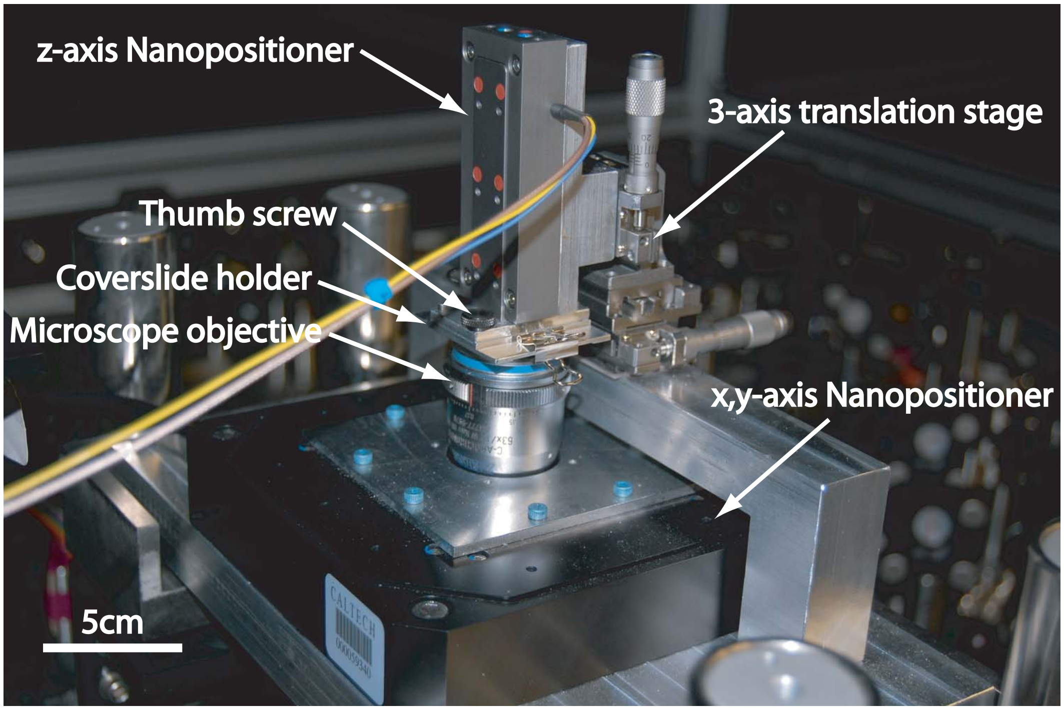


Figure 5.7: Assembled microscope stage. Labeled components are described in the text.

Chroma Technology and a set of dichroic mirrors from Chroma Technology, Omega Optical, and CVI Laser. A schematic of the filter arrangement is shown in Fig. 5.6. The Chroma bandpass filters have very high passband transmission (> 90%) and stopband optical densities greater than OD6. Over time we have slowly replaced all of our bandpass filters with Chroma's ET and HQ series filters because of their exceptional performance. The dichroic mirrors all perform well, with seemingly negligible differences between manufacturers. Of note, however is that the mirror by CVI Laser was mistakenly used in making the measurements found in Chapter 6, and reduced the detection efficiency (for fluorescence at 655nm) by 80-90%.

5.2.4 Mechanical components

Sample stage.

A stable mechanical stage is necessary to hold the nanopositioners, sample, and microscope objective in the proper relative orientation. I designed a fairly simple stage to accomplish this and fabricated most of the necessary parts out of aluminum in the Caltech mechanical engineering machine shop. The assembled stage is shown in Fig. 5.7.

The xy nanopositioner is bolted to a 1.5" thick aluminum plate that is suspended above the optical table on rubber balls for vibration reduction. A 1/8" thick aluminum plate

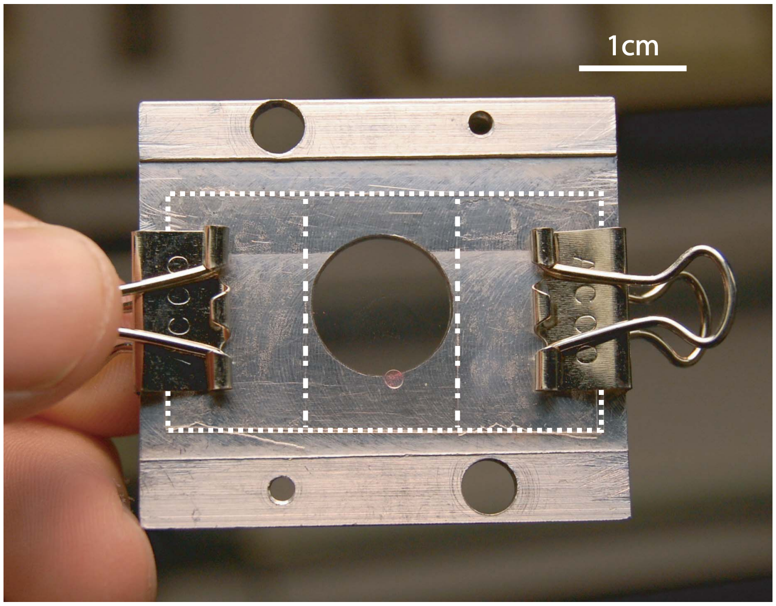


Figure 5.8: Coverslide holder. White dashed rectangle indicates the border of the coverslides. White dot-dash lines indicate the sample border imposed by the double-sided tape. The 12.5mm central hole reduces laser back-reflection. 5mm corner holes are for dowel pin alignment and the smallest holes are threaded for fastening with a thumb screw.

and some Thorlabs mounting hardware hold the microscope objective in place atop the nanopositioner. This configuration is stable enough for the microscope objective to be removed and replaced without requiring realignment of the laser beams.

25μL liquid samples are sandwiched between two VWR No. 1 glass microscope coverslides that are separated by a layer of double-sided tape to ensure sufficient sample depth (about 100μm). A pair of binder clips clamps the coverslides to an aluminum holder that attaches firmly to the z nanopositioner via a complementary aluminum plate. The coverslide holder is held in place by a single thumb screw, and held in alignment by a pair of dowel pins. This mechanism was efficient for loading and unloading samples quickly, and held them in place quite stably. The coverslide holder with slides and binder clips are shown in Fig. 5.8.

The z nanopositioner is suspended above the microscope objective by a Newport miniature 3-axis translation stage used for coarse adjustment. This stage was a disappointment because of its abominable torsional rigidity. Fortunately, there is no evidence that its instability affected the performance of the apparatus, which perhaps should not be surprising considering the tiny forces exerted and distances traveled by the nanopositioners.

Nanopositioners. Both of our nanopositioner stages came upon us by a bit of serendipity. The xy nanopositioner (Physik Instrumente P-517, 100x100μm) was purchased for scanning the surfaces of coverslides, before a tracking apparatus was envisioned, but happened to perform well enough to track fairly fast-moving particles. The z axis nanopositioner (Physik Instrumente P-753, 38μm) was purchased long ago for another fluorescence experiment, and I happened to discover it in a closet while looking for some glassware.

The two-axis stage has an unloaded resonance at about 400Hz, with -90° of phase at 300Hz. Adding the water immersion microscope objective and mounting bracket brings the resonance down to 300Hz, with -90° of phase at about 220Hz. These resonances are unavoidable consequences of using long-travel piezo stages and do limit our tracking bandwidth. In contrast, the z-axis stage has both loaded and unloaded resonances at over 1kHz due to its much shorter travel and very light-weight load. This stage is fast enough to not limit our tracking bandwidth in any detectable way, but its shorter travel means that it is usually the limiting factor determining the duration of tracking trajectories. It may eventually need to be replaced by, or used in tandem with, a stage with longer travel.

5.2.5 Acquisition hardware and software

Avalanche photodiodes. We use industry-standard Perkin Elmer photon-counting actively-quenched APDs (SPCM-AQR15). These output 35ns TTL pulses each time a photon is detected, with a 50-60ns dead time following each pulse. Being careful to shield the detectors from ambient light, it is possible to operate them very nearly at their dark count rates provided the excitation power is not high enough (say, 10μW) to induce significant autofluorescence from the glass coverslides or leakage of the excitation beam through the emitter filters. The dark count rates of these detectors decrease after initial power-up to a steady-state minimum after about 1 hour of warming up. In addition, I have found evidence that the dark count rates on these detectors *increase* over much longer timescales. When I began testing, the detector that we had used most frequently had a dark count rate roughly 12 times higher (600Hz) than the manufacturer specification (50Hz). Our other, less-used detectors all had count rates only twice as high as the specification. The higher dark-count detector eventually stopped working completely, so it is possible that this is a symptom of impending failure.

Time Interval Analyzer. We use the GuideTech GT653 time interval analyzer (TIA) board to record photon arrival times with 75ps (same-channel) and 500ps (cross-channel) timing accuracy. Using the API provided by its manufacturer I wrote code in C++ to interface the boards with Matlab, and this interface has worked fairly well. I am able to record data essentially indefinitely using the board's memory-wrap function, although I usually limit acquisitions to about 100 seconds. Since upgrading to a new version of MATLAB a memory allocation error has begun to appear about once a week, requiring a re-start of MATLAB each time. I am confident that my code does not have memory leaks or overrun errors; this may be a problem with using executables compiled against an earlier version of MATLAB's MEX libraries, but I have not taken the time to recompile the executables yet.

Computer data acquisition board. A National Instruments data acquisition board (NI-DAQ) records analog signals from our apparatus. The 2-D tracking apparatus had used two separate computers for the NI-DAQ and GT653 boards because of system failures that had occurred with early versions of the National Instruments driver software. I reverted to using a single computer for all acquisitions after installing upgraded drivers that seemed more stable, and I began using the MATLAB Data Acquisition Toolbox to control the NI-DAQ. This was convenient, but had a bug that I never quite tracked down: when the computer was recording photon counts at a high rate (several hundred kHz), the sampling on the NI-DAQ would slow down without triggering an error in the software. As a result I obtained data traces reporting that, for example, the fluorescence signal from a single tracked particle spanned 30 seconds while the stage motion from the same particle spanned 40 seconds. This was never really a problem because I rarely operated at such high count rates, but it is something that should not happen.

5.2.6 Control Electronics

Electronics — most home-made — are responsible for synchronizing the beam modulation and controlling the tracking stages and excitation beam power. They can be broadly divided into two categories: modulation electronics and post-detection electronics. Figures 5.10 and 5.11 are electrical schematics of the modulation and post-detection electronics, respectively. Most of these electronics are visible in the photo in Fig. 5.9.

5.2.6.1 Modulation electronics

VCO controller. This circuit takes the sinusoidal reference signal from the xy lock-in amplifier and transforms it so that it may be used to drive the VCOs for beam rotation. A basic schematic is shown in Fig. 5.10. The circuit makes three identical copies of the input reference signal and applies an independently adjustable gain (for rotation radius control) and a single adjustable DC offset of about 9V (to center the VCO at the AOM's peak response frequency) to each one. One of those signals is output to the y-axis VCO. In order to ensure that the modulations along the x and y axes have the necessary 90° phase difference, the remaining two signals go through all-pass filters that apply variable phase shifts[55] before being output to the x-axis VCOs. The phase must be set any time

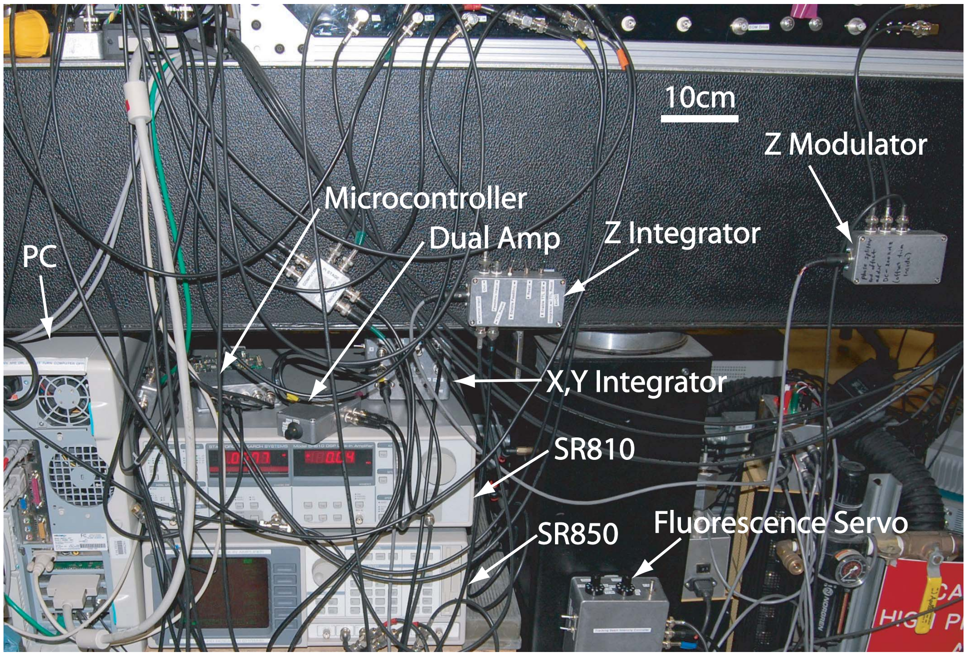


Figure 5.9: Photo of most of the electronic components of the apparatus. Cable connections at the top right corner are part of a light-proof box containing the optics, detectors, and sample stage. Labeled components are all described in the text.

we change the rotation frequency, because both the filters' phase shifts and the phase lags due to the acoustic wave propagation delay in the AOMs depend on the frequency.

Z Modulator. This circuit is a simple op-amp phase splitter[55]. A basic schematic is shown in Fig. 5.10. It takes as its input the reference signal from the z-axis lock-in amplifier, splits it in two, and inverts one of the resulting signals. An adjustable DC offset (of about 0.3V) is then applied to each signal, and the signals are output to mixers that modulate the power in the two tracking beams. By carefully aligning the beams into the two AOMs, the acoustic wave propagation delays can be made almost identical for the two beams. This means that we may come very close to achieving the 180° relative phase between the two beams without using variable phase shifters. Our apparatus actually has a phase error of about 5° between the two beams, but the magnitude of the localization signal varies quadratically with small phase errors so that this contributes only about a 0.2% loss in signal.

5.2.6.2 Post-detection electronics

Pulse splitting and combining electronics. The signals output by the APDs are sequences of 35ns pulses. These pulses must be recorded by a computer, made available for real-time monitoring, and processed by the lock-in amplifiers to compute localization signals.

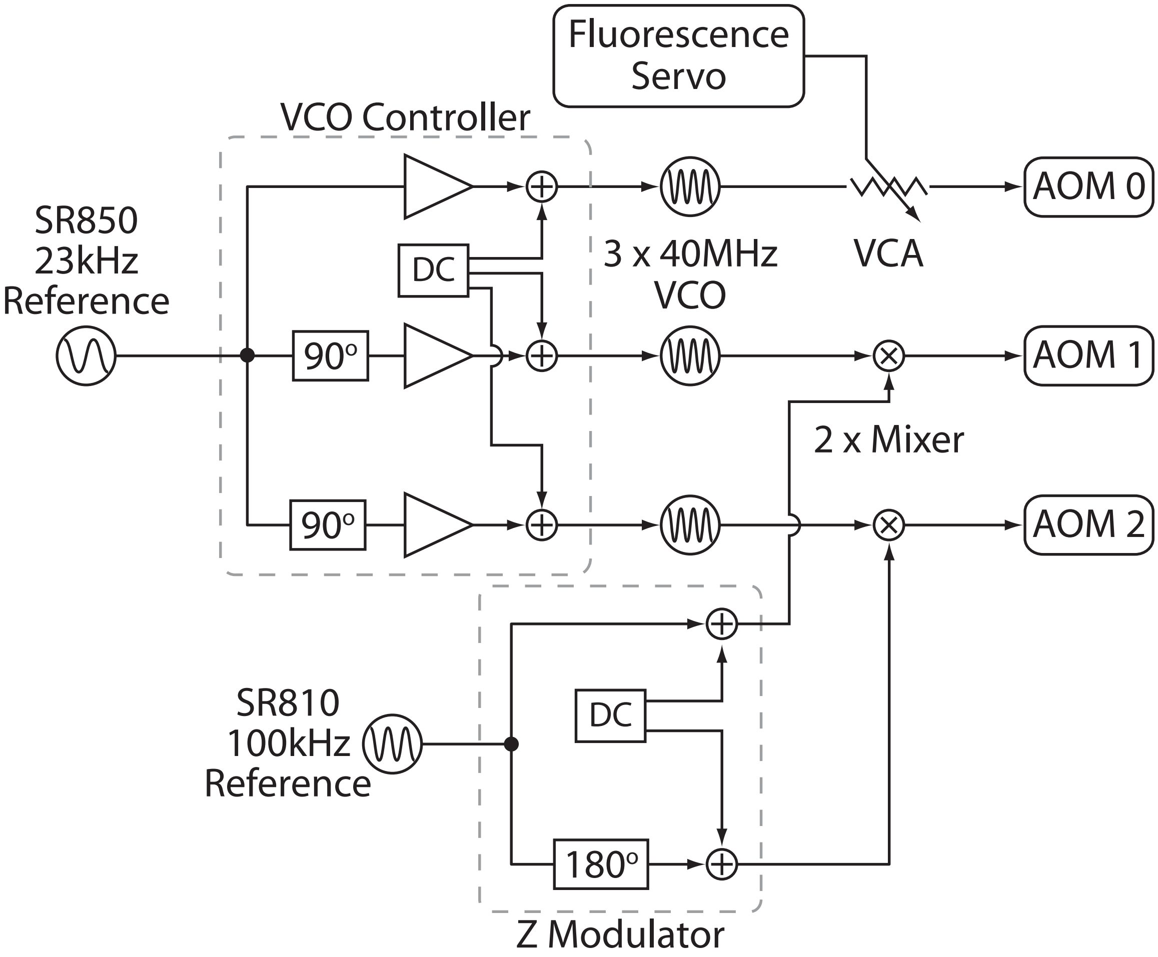


Figure 5.10: Schematic of the electronics for beam modulation. Components grouped in dashed rectangles comprise individual home-built circuits and are described in the text. The home-built Fluorescence Servo appears in Fig. 5.11. The remaining components are commercial items. VCO: voltage-controlled oscillator, Mini-Circuits ZOS-50. Mixer: Mini-circuits ZAD-8. VCA: voltage-controlled attenuator, Mini-Circuits ZX73-2500.

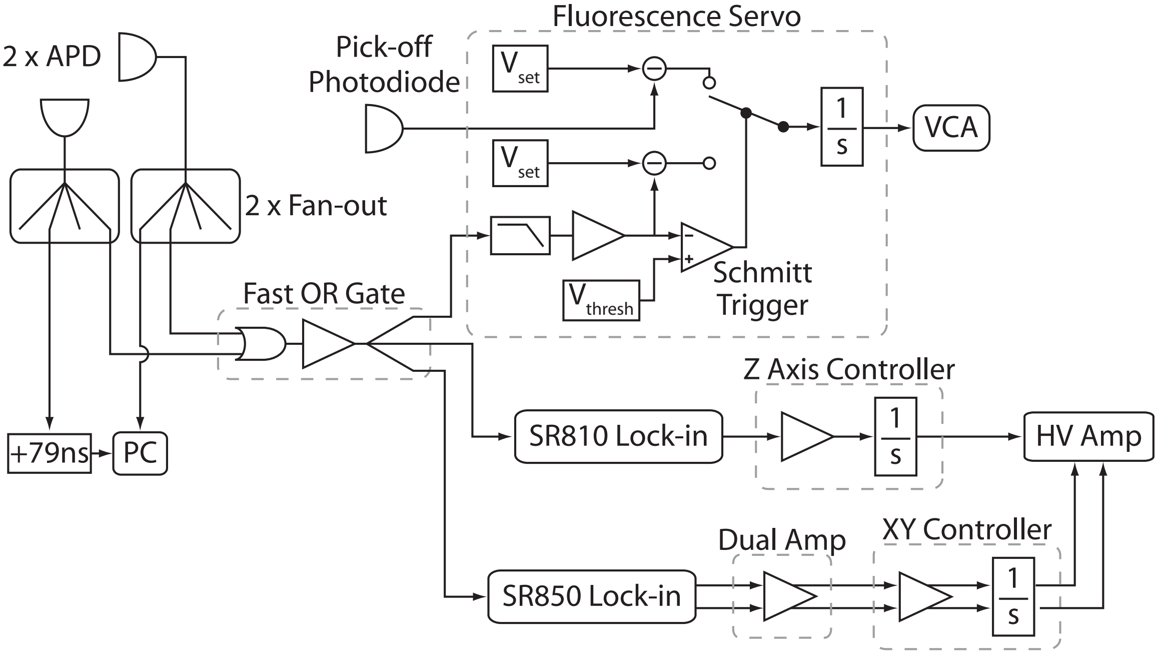


Figure 5.11: Schematic of the detection and feedback electronics. Components grouped in dashed rectangles comprise individual home-built circuits and are described in the text.

Figure 5.11 contains a schematic of the electronics used to split and combine the signals as necessary. A pair of commercial TTL fan-out circuits (Pulse Research Labs) outputs four copies of each APD signal. One copy is recorded by the GT653 TIA in the computer, with a 79ns analog delay imposed between channels by a long coaxial cable. This delay reduces noise at very short lag times due to the coincident arrival of pulses at the TIA. A second copy of the fluorescence signals is filtered and amplified for display on an oscilloscope (not shown). A third copy is combined using a home-built circuit based on a fast TTL OR gate (SN74F32) and buffer (SN74F125). One copy of the combined signal is output directly to each lock-in amplifier, and one copy is output to the fluorescence servo. 50Ω terminators are used to prevent the reflection of the high-frequency components of the signal.

Lock-in amplifiers. We used the dual-phase SR850 DSP lock-in amplifier by Stanford Research Systems for demodulation on the x and y axes for both 2-D and 3-D tracking experiments. For z-axis demodulation I initially built an analog lock-in amplifier using a VCO chip phase locked to the excitation beam power; a mixer for phase detection; and a low-noise amplifier. This circuit worked, and I used it to record my first 3-D tracking data on particles as small as 24nm, but the mixer's temperature-dependent DC offset presented a problem. The offset can only be trimmed when the apparatus is not tracking a particle. However, as soon as the apparatus began tracking a particle and the fluorescence signal jumped up, more power was dissipated in the mixer and it heated up. This caused the offset to drift, introducing a time-dependent bias in the lock point of the tracking system. To avoid such errors, and to take myself out of the feedback loop (as the offset trimming servo), I purchased the SR810 single-phase DSP lock-in amplifier. This change was also the primary reason that I began using the AOMs for the power modulation — the higher frequency used with the EOM was too fast for DSP amplifiers, and I did not want to use an analog lock-in because of their inferior DC offset properties. As a result of these changes, the stability and robustness of the tracking apparatus improved remarkably.

I operate the lock-in amplifiers with an integration time of $300\mu s$ and with the lowest-possible dynamic reserve. This means that the lock-in applies as much gain as possible to the analog input signal, which reduces spurious digitizer noise. After some investigation, I convinced myself that high dynamic reserve settings were a major contributor to the small amount of noise that kept the 2-D tracking apparatus Andy Berglund and I built from reaching the shot-noise limit. We vary the loop gain of the feedback system using the adjustable sensitivity settings of the lock-in amplifiers. The variable aspects of the loop gain are the fluorescence rate and the lock-in sensitivity, so we must balance changes in the fluorescence rate by changing the sensitivity in order to maintain a particular tracking bandwidth. The lock-in sensitivity is adjustable only in fairly large (6-8dB) increments, so I built a variable-gain dual amplifier with calibrated 2dB increments to make more sensitive adjustments on the x and y axes. The z axis was less sensitive to gain changes due to the z nanopositioner's high resonance frequency, so the lock-in's coarse increments were sufficient.

It may seem counter-intuitive to output digital pulses directly into the (intrinsically analog) lock-in amplifiers. In fact, no signal is lost by doing this[37]. In addition, it is an unfortunate misconception in the closed-loop particle tracking community that it is possible to benefit from complicated demodulation schemes that actuate the tracking stages after every detected photon, rather than performing lock-in detection and averaging over an intermediate time-scale. As shown in [19, 36, 37] and Sections 2.2.4 and 2.3.5, photon-counting shot noise distorts localization estimates made with any finite averaging time. Taking that averaging time to zero is worst-case example of this. Such an approach stems from a basic misunderstanding of localization statistics, but is mitigated by similarly-misunderstood dynamic properties of the feedback system. All feedback actuators have some latency, and so if averaging is not performed on the localization estimate it will be performed by the actuator. For a relatively slow actuator such as a piezo stage, the result is hardly noticeable; for a fast electroosmotic trap, measurable displacement of the trapped particle and localized heating can occur[51].

Tracking controllers. I built the z tracking controller as an integrator based on the LT1012 op amp, with a differential input (based on the OP27 op amp) for properly interfacing the controller ground with the lock-in amplifier ground. For my early 3-D tracking measurements, I used the integrator circuit from the 2-D tracking apparatus, which was based on OP27 and had floating inputs. I eventually replaced this circuit with one identical to the z integrator because of its superior noise and offset properties. Each circuit includes an analog switch (ADG417 for z, SW06 for x and y) that allows the integrator to be reset with a TTL pulse.

High voltage amplifier. We drive the nanopositioner stages with a high-voltage amplifier from Physik Instrumente that features an integrated readout circuit for measuring the stage positions from their capacitive sensors. The outputs of the amplifier have low-pass filters at frequencies of several hundred Hz, and these cause substantial amounts of phase loss within the tracking bandwidth. The roll-off is presumably designed to reduce the gain at the stages' first resonance frequencies, but we would prefer a flat response that allowed us more control over the stage dynamics.

Fluorescence Servo. Any variations in the fluorescence rate of a tracked particle translate into variations in the loop gain of the feedback system. This can cause problems if the variations are too large. If the fluorescence rate becomes too high, high loop gain may cause the feedback system to become unstable; if the rate becomes too low, the tracking bandwidth may become too small and the particle may be lost. When we were working with mixtures of fluorescent beads of different sizes (and different brightnesses), Andy Berglund and I devised a feedback circuit that used an AOM to vary the DC optical power in the tracking beam in order to keep the loop gain constant. This allowed our 2-D tracking apparatus to track 60nm and 210nm beads (with 27-fold differences in brightness) within a single sample, and with exactly the same fluorescence rate[19, 42].

The fluorescence servo that we used was a very simple circuit. A set point was chosen for the fluorescence rate, and the difference between the detected fluorescence rate and the set point was sent to an integrating controller with a bandwidth of a few Hz. When no particle was being tracked, the integrator was fixed on its upper rail so that the excitation beam was as bright as possible. Once a particle drifted into focus, the fluorescence rate quickly jumped above the set point and the integrator reduced the power to bring the fluorescence back down.

We required a fluorescence servo for our work with DNA molecules in Chapter 7 because those molecules were labeled with dyes that photobleach. This approximately exponential decay in the fluorescence rate requires an exponentially-increasing excitation intensity to counteract. As a result, the amount of time for which we can track a molecule before all its dyes have bleached is proportional to the logarithm of the maximum amount of optical power available to the fluorescence servo. Consequently, we want the fluorescence servo to have as much available power as possible. However, the simple fluorescence servo allows the excitation beam to quickly reach its maximum intensity while no molecules are being tracked. When a molecule finally does drift into focus, this high intensity will quickly destroy many of its dyes because the bleaching rate is proportional to the intensity. Even in the short period of time between the molecule drifting into focus and the fluorescence servo reducing the beam intensity, a lot of damage will be done.

I developed a more sophisticated version of the fluorescence servo that works around this problem. The servo inputs both the fluorescence rate and the optical power in the tracking beam. When the fluorescence rate is low (because no fluorescent particles are in focus), the servo locks the optical power to a set point that is usually an order of magnitude or more below the total available optical power. Once a molecule drifts into focus and the fluorescence rate increases, the servo transitions to locking the fluorescence rate to a separate set point. As the molecule bleaches, the servo is able to increase the excitation intensity by a very large amount, but the optical power when the molecule first drifted into focus is completely unrelated to the size of this power reserve.

A schematic of the improved fluorescence servo appears in Fig. 5.11. A Schmitt trigger[55] (based on the LT1677 op amp) is used to discriminate between low and high fluorescence values — this is a comparator that triggers on and off at two adjustable threshold values. This hysteresis is necessary because of the large amount of photon counting noise at the low fluorescence rates that we trigger on — we do not want counting noise fluctuations to cause errant switching between the servo's operational modes. A single analog integrator (based on the LT1012 op amp) is the only controller in the servo in either operational mode. When the Schmitt trigger makes its transitions, it simply switches (via an ADG436 analog switch) the integrator's input between the fluorescence error signal and the optical power error signal. The integrator outputs to a voltage-controlled attenuator that regulates the RF power into the y-axis AOM.

Analog microcontroller. Whenever a particle is lost at the boundary of one of the tracking stages, it is best to return the stages to the center of their ranges in order to maximize the duration that subsequent particles are tracked for. We use an analog microcontroller (Analog Devices ADuC812) to reset the integrators when they are near their boundaries. I wrote code in assembly that does this somewhat badly, as it allows the z stage to rail on its upper boundary.

A microcontroller was used in the past as the feedback controller in the initial 2-D tracking apparatus, but its output was noisy so we switched to an analog controller. I eventually tracked the microcontroller noise down to a grounding problem that I remedied by putting a differential amplifier on the DAC output. The noise was still worse than an analog circuit, and I never did find a good application for the microcontroller, but it will be good to know this if one should arise in the future.

5.3 Experimental procedures

In this section we describe certain procedures that were necessary for constructing or calibrating the tracking apparatus.

5.3.1 Bead immobilization

Immobilized fluorescent beads are extremely useful diagnostic tools because they facilitate direct characterization of the focused tracking beams. In any tracking application this is important for ensuring the quality of the modulated beams and the alignment of the detectors[37]. For our three-dimensional tracking apparatus, it is no understatement to say that the ability to produce immobilized fluorescent beads reproducibly is essential.

We discovered by accident that coating glass coverslides with a layer of poly-l-lysine causes strong adhesion of polystyrene beads. At the time we were trying to find a way to *prevent* positively-charged amino-derivatized quantum dots from sticking to glass; we simultaneously failed at our goal and made one of the most important strides in all of the years of the development of our apparatus. In particular, the 210nm fluorescent polystyrene beads with “Suncoast Yellow” dye made by Bangs Laboratories stick instantly and irreversibly to these charged surfaces. These beads are inexpensive, extremely photostable, extremely bright, and absorb light with wavelengths spanning from the UV up to nearly 650nm.

For cleaning and coating coverslides, I fabricated a pair of teflon racks that hold about 20 coverslides each. The racks are held together by steel screws and brass thumb nuts that discolor quickly in corrosive cleaning solutions, but this does not seem to affect the slide coating. The procedure we used to coat the coverslides with poly-l-lysine was a very slight variation (substitution of KOH for NaOH) on a procedure made available online by Pat Brown’s lab at Stanford[56]. Once the slides have been coated, they may be used essentially immediately to immobilize beads. We dilute the stock 1% bead solution by a factor of about 10^{-3} into a 0.1% v/v solution of Tween-20 (Sigma-Aldrich P2287, a detergent that reduces bead aggregation) in purified water. We generally prepare about 1mL of this solution at a time because it is

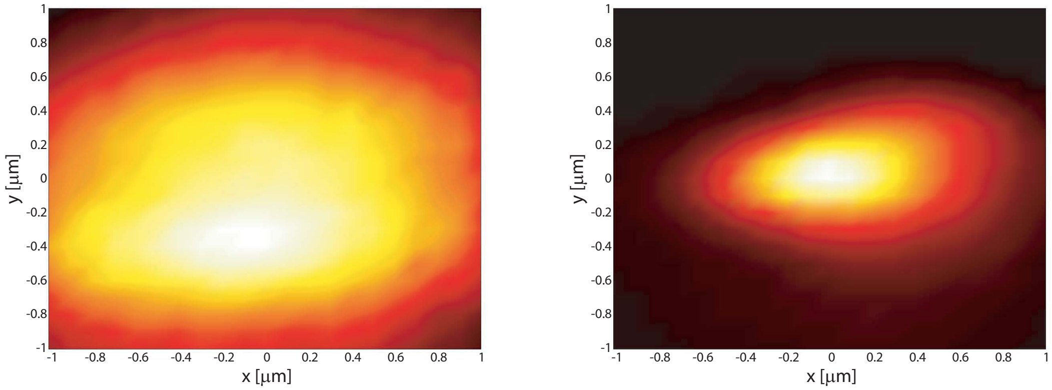


Figure 5.12: Position-dependent fluorescence intensity of the rotating 532nm tracking beam (left) and the 415nm probe beam (right) measured on an immobilized 210nm fluorescent bead. Raster scans used 100nm increments along both axes. We fitted Gaussians to both beams and corrected for the 210nm bead size. Corrected waist fits: 532nm: w_x = 2540nm, w_y = 1790nm. 415nm: w_x = 710nm, w_y = 375nm.

stable for years without precipitation or photodegradation even when exposed to room lighting. We sandwich 2μL of the diluted bead solution between a poly-l-lysine slide and an uncoated glass slide, and the beads usually adhere in less time than it takes to place the slides in the microscope.

5.3.2 Beam scans

Using an immobilized bead we can map out — in three dimensions if we choose — the spatially-dependent excitation intensity of each beam. The capacitive sensors in the nanopositioners make this easy — an internal feedback servo in the high voltage amplifier can be used to position the stages in closed loop with sub-nanometer accuracy. We use the NI-DAQ to raster scan the beam across an immobilized bead, and we average the fluorescence at each point for a few hundred milliseconds using the TIA board. The resulting data can be very helpful for estimating the waists of the tracking beams, from which we can then infer the statistics of the localization estimates. This type of scan can be useful to reveal asymmetries or other irregularities in the beams.

Figure 5.12 shows the focused beam spots for the rotating 532nm tracking beam and a 415nm probe beam, measured on an immobilized fluorescent sphere. The tracking beam was particularly large when these measurements were made — it may not have been properly focused — so the large ratio of beam sizes is a bit skewed from normal. However, the waists of the 415nm beam are almost certainly too small to be useful for tracking anything but very slow-moving particles. The scan of the 415nm beam indicates a slight astigmatism — the beam focuses along the x and y axes in different planes, resulting in elliptical cross-sections. We saw very similar astigmatism when we focused the 532nm beam to a diffraction-limited spot, implying that it may be caused by distortion in an optical element shared by the two beam paths.

We must be cautious in our use of beam scans to characterize the probe and tracking beams quantitatively. Unlike our 2-D tracking apparatus, in which samples were only a few μm thick, 3-D tracking requires the transmission of the beams deep ($\sim 50\mu$m) into a liquid sample. Measurements that we make on beams at the glass surfaces may be significantly distorted by optical aberrations when the beams are used for tracking. An extreme example of this occurs when we use an oil immersion microscope objective: the oil has a refractive index that is matched to the glass but severely mismatched to the water. The deeper the beams are focused into water, the more distorted they become. We were not able to track in three dimensions using an oil immersion objective. Our water immersion objective is designed to minimize distortion, but we are not able to know with certainty how effective it is. Open-loop FCS measurements (see Section 4.1) yield numbers that are consistent with our measured beam shapes, but these measurements are hardly sensitive enough for making careful comparisons.

5.3.3 Localization signal measurement

Another important application of an immobilized fluorescent bead is the measurement of localization error signals. The quality of these signals determines the tracking fidelity of the apparatus, so the ability to directly measure them is essential for proper alignment of the tracking beams. We use the nanopositioners to scan the bead through the center of the modulated laser beams, and we simultaneously record the stage positions and the output of the lock-in amplifiers. Figure 5.13 shows the localization signals that we measured immediately before capturing our first-ever three-dimensional tracking results.

5.3.4 Alignment

The alignment of the tracking beams relative to each other is important to ensure quality localization signals. As we began discussing in the previous section, the beam alignment may be characterized using the localization error signals measured on an immobilized fluorescent bead. In this section we describe in more detail the characteristics of a well-aligned pair of beams.

The two focused tracking beams must be both collinear and normal to the coverslides. This can be ensured by satisfying two conditions: the light back-reflected from the coverslides by both beams should produce concentric ring diffraction patterns that expand and contract symmetrically with coarse changes in the z position of the sample, and the two-dimensional

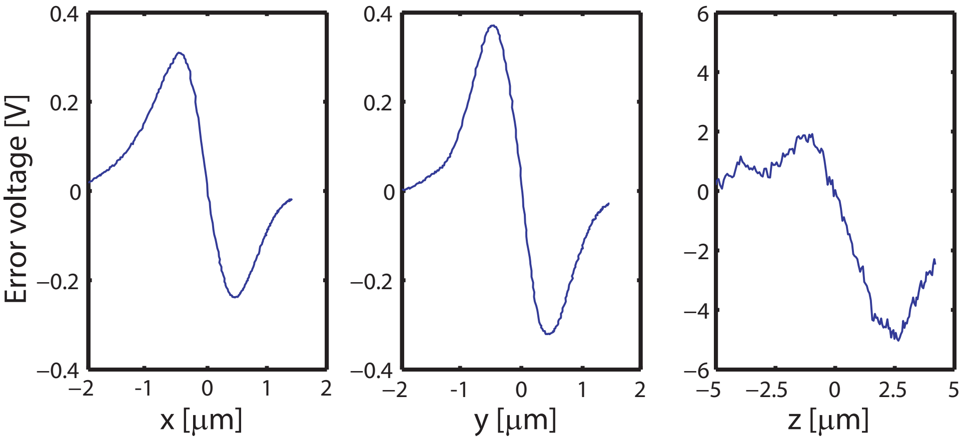
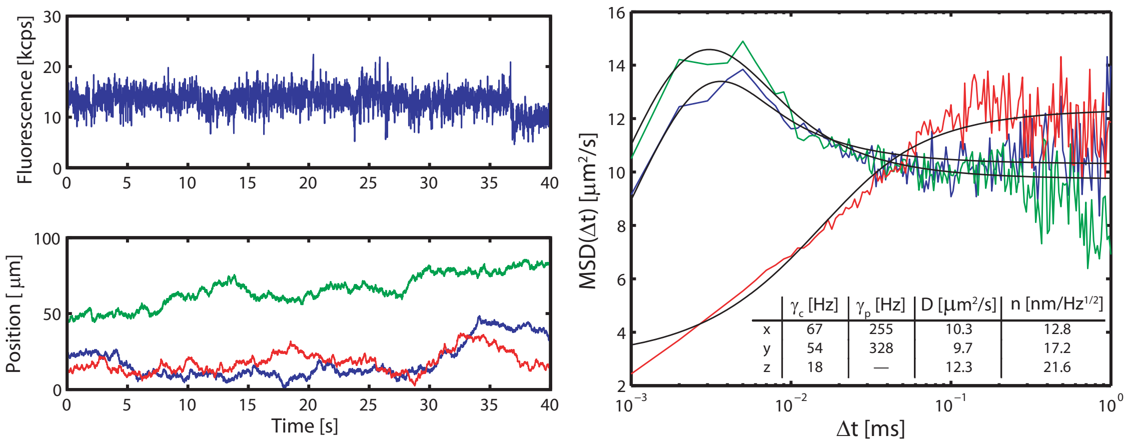


Figure 5.13: Localization error signals measured on an immobilized 210nm fluorescent bead. Each data point is an 80μs average of the analog signal. The x and y signals were computed by the SR850. The z signal was measured with a home-built analog lock-in amplifier — its higher intermediate frequency accounts for the larger noise on that signal. Peak fluorescence rates were several hundred kHz. Signals were recorded on a Tektronix TDS3032 oscilloscope.

localization signals from both beams should go to zero at the same point when scanned over an immobilized fluorescent bead. With care, it is possible to eliminate all measurable deviations from this ideal. However, the alignment tends to drift over time, probably due to some combination of thermal expansion of mirror mounts, slow AOM pointing drifts, and accumulation of slight displacements of the tracking stage with use. The effects of drift are particularly pronounced on the 415nm probe beam because of its small waist. This beam can easily be deflected by an entire beam waist but, unlike tracking beam errors that can be difficult to diagnose, such a deflection produces obvious results. With regular use, the alignment of all beams should be checked every week or two to ensure good tracking system performance.

The criteria we specified for good beam alignment along the x and y axes never really fails. Unfortunately, no comparably reliable criteria exist to ensure good z-axis alignment. Due to the extremely small offset ($\sim 1\mu$m) desired between the two tracking beams relative to the focal length of the microscope objective ($\sim$ 1mm), it is difficult to try to prepare beams with the necessary divergences simply by measuring the propagation parameters of the beams. Things are certainly not simplified by the fact that our tracking beams have the asymmetric geometry described in Section 2.3.1, and the rotation radius therefore factors into the location of the tracking plane and the quality of the localization signal. Our only option tends to be to make careful adjustments of the beam divergence that we monitor directly on an immobilized fluorescent bead. In general, a well aligned pair of beams with properly-balanced optical powers will produce a nearly symmetric z-axis localization signal that goes to zero when the beams are positioned so that the brightest point of their combined excitation profile lies over an immobilized fluorescent bead. Exactly how to achieve this ideal, and how deviations from it



	γ_c [Hz]	γ_p [Hz]	D [$\mu m^2/s$]	n [nm/$Hz^{1/2}$]
x	67	255	10.3	12.8
y	54	328	9.7	17.2
z	18	—	12.3	21.6

Figure 5.14: **Left**: Fluorescence rate and positions of the x (blue), y (green) and z (red) stages while tracking a 24nm fluorescent bead. **Right**: Mean-squared deviation of the tracking stages as a function of time interval Δt. Second-order fits to Eq. 3.35 are shown in black for x and y, and a first-order fit to Eq. 3.34 is shown for z. Fit parameters are given in the inset table.

affect tracking performance, is not always apparent. Often, a lot of trial and error are needed; probably the only sure way to optimize the z-axis alignment is through patience and careful diagnostic testing. Fortunately, unlike the x and y axes the z-axis alignment tends not to drift at all. Once the proper alignment is found, it may never require adjustment again.

5.3.5 Tracking system characterization

Once the beams are aligned and producing good error signals, freely-diffusing fluorescent beads take over the diagnostic role of their immobilized counterparts. A particularly useful measurement is made by tracking a bead and computing the feedback system parameters described in Section 3.2.2 from the function $MSD(\Delta t)$ defined in Eq. 3.34. By measuring these parameters we can set the electronic gains in the feedback loops to appropriate values for particular fluorescence rates.

Figure 5.14 shows some very early tracking data, taken seven days after the initial success of the three-dimensional tracking apparatus. The sample consisted of a dilute solution of 24nm fluorescent beads (Invitrogen FluoroSpheres) in 0.1% v/v Tween-20 in water. We sonicated the mixture for an hour to reduce bead aggregation, but we still see evidence of bead clusters in solution. We fitted curves from Eq. 3.35 to the mean-squared deviation data and found a diffusion coefficient $D = 10.8 \pm 1.6\mu m^2/s$, consistent with a cluster of beads approximately 40nm in diameter. The fluorescence rate shown in the figure fluctuates about its mean value because of tracking errors, some caused by systematic drifts in the analog lock-in amplifier used on the z-axis. The abrupt drop in fluorescence at about $t = 37s$ may be due to one of the beads breaking away from the cluster, but we do not have enough data to confirm that the

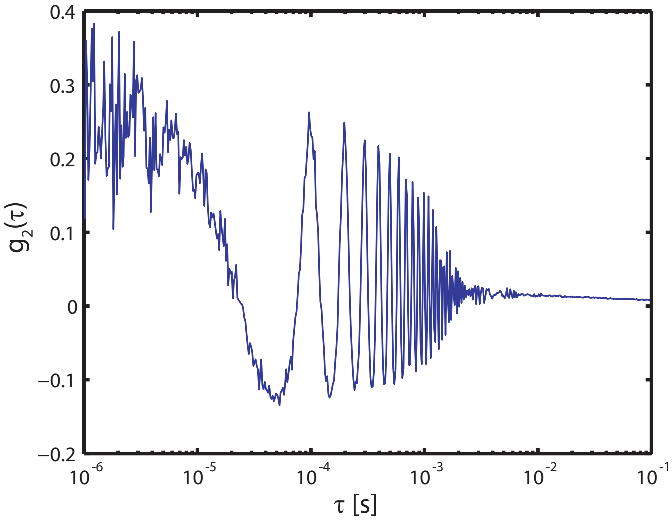


Figure 5.15: Tracking-FCS curve for 24nm fluorescent bead data in Fig. 5.14. 10kHz beam rotation accounts for the large oscillations. 338kHz z-axis power modulation is much more subtle.

diffusion coefficient of the cluster increased after the drop.

We calculate the steady-state tracking error according to Eq. 3.31 and get $L_x = 260$nm, $L_y = 290$nm and $L_z = 370$nm. The loop gain was above its optimal value $\gamma_c = \sqrt{2D/n^2}$ (see Eq. 3.32) on both the x and y axes, and this accounts for the fairly large overshoot at $\Delta t \approx 3$ms in the MSD curves. Reducing the gain would have reduced the tracking error. Likewise, γ_c was below optimal along the z axis so increasing the z gain would have improved the results. Nonetheless, these values compare compare reasonably to the localizations we achieved with the two-dimensional tracking apparatus on slower particles but with fewer photons[19].

Figure 5.15 shows $g_2(\tau)$ computed for the fluorescence measured in Fig. 5.14. As discussed in Section 4.2.4, it is often difficult to extract much information about the z axis beam modulation because of the gradual decay in the fluorescence intensity along that axis. For the data in the figure, the z-axis beam modulation frequency was $\omega_z = 338$kHz and the beams were rotated at 10kHz. A generous eye might be willing to attribute the jagged appearance of $g_2(\tau)$ between 1 and 50μs, or the slight shoulders about the peak at 100μs, to the z-axis modulation. However, there is too much noise at those short times and the shoulders are sensitive to beam misalignment, so it is not clear whether we can resolve the modulation at all.

5.3.6 Delay calibration

For measurements made with a pair of detectors in the Hanbury-Brown-Twiss configuration shown in Fig. 5.4, the inevitable delay δ caused by optical and electronic path length differences has the effect of shifting the location of coincident photon detections in the FCS curve from

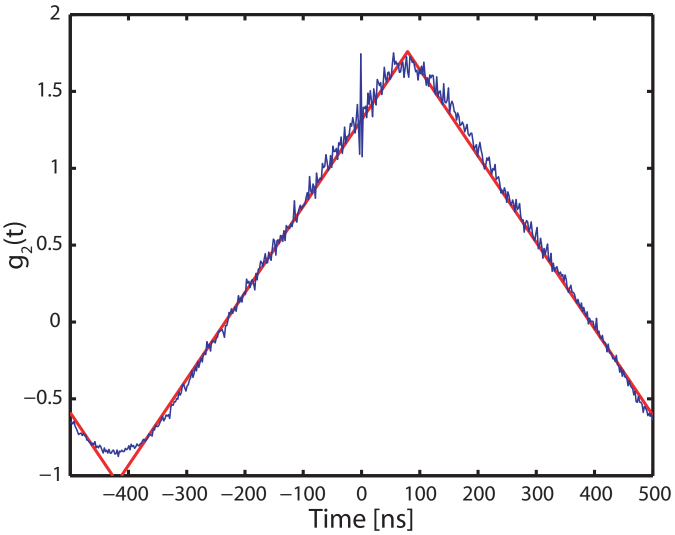


Figure 5.16: $g_2(\tau)$ (blue) measured by strobing the optical power at 1MHz over an immobilized fluorescent bead. Fit (red) to a triangle wave yields a delay $\delta = 79$ns.

$g_2(0)$ to $g_2(\delta)$. We intentionally induce a large electronic delay (see Fig. 5.11) to reduce timing noise at short correlation times (visible at $\tau = 0$ in the figure), and this shifts the FCS curve further. In order to extract information about any processes affecting the fluorescence on timescales that are at all comparable to δ, we must make an accurate estimate of δ and correct our FCS curves by offsetting the delay.

We made our delay calibration for the measurements in Chapter 6 by using a high-speed TTL RF switch (Mini-Circuits ZYSWA-2-50DR) to strobe the RF power in the y axis AOM, and consequently the optical power in the beam, at 1MHz. We approximate the fluorescence intensity with a square wave of period T (= 1μs),

$$I(t) = \begin{cases} 1 : 0 \leq t \pmod{T} \leq T/2 \\ 0 : t \pmod{T} \geq T/2 \end{cases}, \tag{5.1}$$

in which case $g_2(\tau)$ is a triangle wave with a peak at $\tau = 0$. We measured fluorescence for 30s on an immobilized 210nm fluorescent bead with the optical power set to provide an average total count rate of about 320kHz. The resulting $g_2(\tau)$ curve is shown in Fig. 5.16, with a numerical fit to a triangle wave. The fitted wave has a peak at 79ns, which is equal to the delay time δ. Visible in the plot is the jitter at $\tau = 0$ due to cross-talk in the TIA board at short time delays; this is the motivation for imposing the electronic delay between channels.

Chapter 6

Photon statistics of quantum dot fluorescence

6.1 Introduction

6.1.1 Quantum dots

Quantum mechanics describes the behavior of matter on very small spatial scales[57, e.g.]. It predicts that the confinement of a particle to a small region places restrictions on the energies that it may have. In particular, the continuous energy scale predicted by classical mechanics becomes a set of discrete energy levels that are observable when a particle is confined on scales usually well below 1nm. The best-known example of this is the hydrogen atom, in which the positively-charged proton and negatively-charged electron are confined due to electrostatic attraction. The resulting quantized energy levels produce the discrete lines in the sun's emission spectrum that were among the most important experimental motivations to the development of the quantum theory.

Quantum effects are not limited to electrostatically-confined free electrons and protons. Electrons within solids are confined by the boundaries of the solids, but for macroscopic solid particles this confinement is very weak. As a result, the electron energy levels are very close together and their discrete nature cannot be distinguished from a continuum. If a solid particle is made small enough, however, the confinement of the electrons can have a large influence on many properties of the solid. For semiconductor materials, these effects are particularly strong and begin to appear at relatively large particle sizes (1-10nm) that are well within reach of modern synthesis and characterization methods. For a review of the size-dependent properties of semiconductors on these size scales, see [58] or [59].

Quantum dots (qdots) are semiconductor crystals between about 1 and 10nm in diameter. The confined electrons in these crystals occupy discrete energy levels; the energy spacing

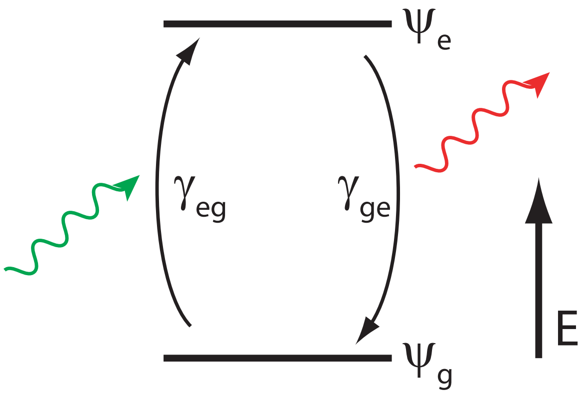


Figure 6.1: Two-level emitter energy level diagram. Absorption of a green excitation photon drives the emitter from the ground state to the excited state. Relaxation to the ground state produces a red fluorescence photon.

between the levels increases, and the strength of transitions between all but a few levels decreases, as the crystals are made smaller[59]. As a result, a simple model of qdots as systems with two tunable energy levels is sufficient to describe some of their properties. In the case of CdSe qdots, the energy level spacing can be tuned to cover nearly the entire visible range of the electromagnetic spectrum, so the qdots become fluorescent emitters of visible light. Figure 6.1 illustrates fluorescence from a two-level emitter. The emitter absorbs incident light at a rate

$$\gamma_{eg} = \sigma_a \frac{\lambda}{hc} I, \tag{6.1}$$

where hc/λ is the energy of the incident photons, σ_a is the *absorption cross section* of the qdot at the wavelength λ, and I is the intensity of the incident light. Absorption of an incident photon induces the transition of an electron within the qdot from its ground (ψ_g) to excited (ψ_e) state. The electron then spontaneously relaxes back to its ground state at a rate γ_{ge} intrinsic to the emitter either by a radiative decay producing a fluorescence photon or a thermal decay producing phonons in the crystal bulk.

Quantum dots were discovered independently by Efros and Ekimov[60, 61] and Brus[62] in the early 1980s. Since then methods have been developed for producing nearly monodisperse[63] CdSe qdots with high fluorescence quantum yields[64, 65], and for conjugating such qdots to biological molecules[66, 67]. Qdots have been shown to be strong 2-photon absorbers; combined with their broad absorption spectra, narrow emission spectra and low biotoxicity, this makes them excellent markers for multi-color microscopy in living cells[58, 68, 69]. In addition, qdots are much brighter and more photostable than organic dyes[69] and their relatively long fluorescence lifetime $\tau_f \equiv \gamma_{ge}^{-1} \sim 25$ns allows qdot fluorescence to be separated from shorter-lifetime background fluorescence using time-correlated methods[70]. For these reasons, qdots today are among the most popular fluorescence markers in biological and single-molecule imaging.

Despite these strengths, however, qdots are not perfect emitters: it is well-known that they exhibit “blinking” behavior, in which their fluorescence turns off and on at random intervals that span a broad range of time-scales[71–74]. Furthermore, a substantial fraction of qdots in any particular sample will generally be permanently nonradiant[75]. Qdots are very sensitive to their local environment [75, 76], which can either complicate or facilitate their use in certain applications. Finally, geometry variations between individual qdots synthesized in the same batch can lead to variation in their fluorescence lifetimes and absorption cross sections[77].

6.1.2 Photon statistics

Earlier in this thesis, we described the light emission from fluorescent particles in very simple terms. A stationary particle in a stationary excitation laser beam radiates photons at a constant rate Γ, with each photon emission independent of each other. In this model the probability $f_1(t)\mathrm{d}t$ of detecting a photon in any time interval $(t, t + \mathrm{d}t)$ is $\Gamma\mathrm{d}t$, and the joint probability $f_2(t_1, t_2)\mathrm{d}t_1\mathrm{d}t_2$ is just $\Gamma^2\mathrm{d}t_1\mathrm{d}t_2$. The correlation function $g_2(\tau)$ of the photon detections from the constant-rate emission model gives $g_2(\tau) = 0$.

The term *photon statistics* refers to the properties of photon detection that are intrinsic to the light-emitting particle. The simple case that we have worked with is known as *Poisson* photon statistics because the probability of detecting n photons in a time interval T is given by the Poisson distribution. While laser light is usually very nearly Poisson[78], most other light-emitting systems exhibit more complicated photon statistics.

An arc lamp is a canonical example of an emitter with *super-Poisson* photon statistics, in which the detection of a photon at a particular time t implies an *increased* probability of detecting a photon at times shortly after t[53, 54]. This occurs due to fluctuations in the electrical current producing the photons: more photons are produced when the current is high and fewer when it is low, so photons are emitted in bunches. This cannot be described in terms of Poisson statistics because the photon detection times are *not* independent of each other.

Light emission from radiative transitions in objects with quantized energy levels produces *sub-Poisson* statistics[79, 80]. Detecting a photon at time t implies a *reduced* probability of detection at times shortly after t because the emitter must make two transitions — from the ground to the excited state and back — for each detected photon. Some time is required to cycle between states, so detection of two photons separated by a time interval shorter than the average cycling time is unlikely. These are known as *anti-bunching* statistics.

6.1.3 Motivation for our work

Quantum dots produce fluorescence light via electronic transitions from the excited to the ground state. As a result, their fluorescence exhibits anti-bunching statistics. Anti-bunching, combined with the blinking statistics described briefly in Section 6.1.1, make for some interesting photon statistics in qdot fluorescence. However, to resolve either requires long observation times: anti-bunching because it occurs on very short time-scales (near $\tau_f \sim 25$ns), so many photons must be detected to develop good statistics; blinking because it occurs over a wide range of time-scales.

It is difficult to measure either process on single qdots in solution because Brownian motion limits qdots' observation times to several tens of milliseconds. Anti-bunching can be resolved if the statistics are averaged over multiple qdots, but variations in the fluorescence lifetime and absorption cross-section are obscured by such measurements. Evidence of blinking shows up as increased variance in the fluorescence signal, but the fluorescence correlation decay is dominated by diffusion so it is difficult to distinguish blinking qdots from objects that do not blink at all[47]. Wide-field microscopy is better at resolving blinking because the observation times of each qdot are much longer, but current detectors do not have sufficient time resolution to detect anti-bunching and an axial tracking system is necessary to keep the qdots from exiting the imaging plane. Consequently, most measurements are made on qdots immobilized on glass surfaces. However, glass surfaces are very reactive[81], and it cannot be known what effect this has on the measured fluorescence statistics.

We used our tracking apparatus to make the first long time-scale high time resolution measurements on qdots diffusing freely in aqueous solution. We resolve photon anti-bunching on individual qdots and find evidence that heterogeneity in the anti-bunching time-scales first observed on glass surfaces[77] was not an artifact of the glass, but persists in solution. The fact that we resolved anti-bunching on single freely-diffusing emitters is itself significant: anti-bunching is the fastest process that can be studied by fluorescence because the emitter cannot respond to any faster fluctuations. In essence, it proves that we are able to resolve *all* dynamic time-scales in the detected fluorescence signal. Ours are the first measurements to achieve emitter-limited time resolution on an individual freely-moving fluorescent object.

In addition to anti-bunching, our measurements are the first to show unambiguous evidence of blinking in qdot FCS curves on time-scales longer than the diffusion time. We used 2-mercaptoethanol, a known blinking suppressant[76], in order to ensure that qdots did not blink off for long periods of time and escape the tracking apparatus. On time-scales longer than 1 second, previous measurements had shown that 2-mercaptoethanol increased the total on-time of blinking qdots by increasing the average time between transitions from the bright to dark states while not affecting the time between transitions from the dark to bright states. Our measurements show that this cannot be the case on shorter time-scales because our data implies that one, if not both, of these time-scales must *decrease* with addition of 2-mercaptoethanol.

This chapter provides background on both anti-bunching and blinking statistics, and presents our measurements of these statistics in qdot fluorescence. It is a more detailed version of the work published in [20].

6.2 Photon statistics of a two-level emitter

Any system with discrete energy levels must be described probabilistically[8]. In the case of the two-level emitter in Fig. 6.1, we define the probabilities ρ_g and ρ_e that the emitter is in the ground and excited states, respectively. γ_{eg} is the rate at which the ground-state emitter is driven to the excited state, and γ_{ge} is the rate at which the excited-state emitter relaxes to the ground state. Given this model, the populations of the two states evolve according to the Master equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{pmatrix}\rho_g^t \\ \rho_e^t\end{pmatrix} = \begin{pmatrix}-\gamma_{eg} & \gamma_{ge} \\ \gamma_{eg} & -\gamma_{ge}\end{pmatrix}\begin{pmatrix}\rho_g \\ \rho_e\end{pmatrix}, \tag{6.2}$$

so that the steady-state populations are given by

$$\rho_g^\infty = \frac{\gamma_{ge}}{\gamma_{eg}+\gamma_{ge}}, \rho_e^\infty = \frac{\gamma_{eg}}{\gamma_{eg}+\gamma_{ge}}. \tag{6.3}$$

We should note that this rate equation model does not account for the deeper physical phenomena governing the qdot's interaction with the excitation laser: in particular, the electric polarization of the qdot oscillates with the electric field of the laser, coupling the qdot to the field. However, strong dephasing effects typical in solid objects, and the broad linewidth of the qdot absorption spectrum, ensure that this oscillation is only relevant on time-scales much shorter than the decay times γ_{ge}^{-1} and γ_{eg}^{-1}[27].

We use the functions f_n defined in Section 4.1.1 to characterize the statistics of the two-level emitter. At any time t, the probability of a photon detection occurring in the time interval $(t, t+\mathrm{d}t)$ depends on the probability ρ_e^t that the emitter is in the excited state and the probability $\gamma_{ge}\mathrm{d}t$ that it relaxes to the ground state in this interval:

$$f_1(t)\mathrm{d}t = \Phi\rho_e^t\gamma_{ge}\,\mathrm{d}t, \tag{6.4}$$

where Φ is the *fluorescence quantum yield*, representing the probability that the $\psi_e \to \psi_g$

transition produces a fluorescence photon. At steady-state, this gives

$$f_1(\infty) = \frac{\Phi\gamma_{eg}\gamma_{ge}}{\gamma_{eg} + \gamma_{ge}}, \tag{6.5}$$

which is the rate of fluorescence photon emissions averaged over all time.

The dynamic behavior of the emitter is contained in the function f_2, which gives way to the correlation function $g_2(\tau)$ as defined in Eq. 4.2. There are several ways to compute f_2: it can be found by evolving the populations ρ_g^t and ρ_e^t according to Eq. 6.2 or by computing waiting time statistics for the times between photon detections. The former approach is less mathematically sophisticated and more generally applicable, but the latter is an interesting physically-motivated approach. I will provide both derivations here.

6.2.1 Evolution of the Master equation

To compute the correlation function we require the two-time detection probability $f_2(t, t+\tau)$, which we can find by conditioning on the first detection time:

$$f_2(t, t+\tau) = f_{1|1}(t+\tau|t) f_1(t), \tag{6.6}$$

where, as defined in Section 4.1.1, $f_{1|1}(t_2|t_1)\mathrm{d}t_1\mathrm{d}t_2$ is the probability of detecting a photon in $(t_2, t_2 + \mathrm{d}t_2)$ *given* that we detected a photon in $(t_1, t_1 + \mathrm{d}t_1)$. If a photon has been detected in $(t, t+\mathrm{d}t)$ then immediately after time t the emitter is, with certainty, in the ground state ($\rho_g = 1$, $\rho_e = 0$). The probability of detecting a photon at a later time $t+\tau$ is then determined by $\rho_e^{t+\tau}$ as it evolves from this initial state. $\rho_e^{t+\tau}$ is given by the solution to Eq. 6.2 with this initial condition,

$$\rho_e^{t+\tau} = \begin{pmatrix} 0 & 1 \end{pmatrix} \exp\left[\begin{pmatrix} -\gamma_{eg} & \gamma_{ge} \\ \gamma_{eg} & -\gamma_{ge} \end{pmatrix}\tau\right] \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \tag{6.7}$$

where the matrix exponential is computed in the standard way by diagonalizing the evolution matrix. After a lot of algebra, we get

$$f_2(t, t+\tau) = \Phi\rho_e^{t+\tau}\gamma_{ge} f_1(t) = \left(\frac{\Phi\gamma_{eg}\gamma_{ge}}{\gamma_{eg} + \gamma_{ge}}\right)^2 \left[1 - e^{-(\gamma_{eg}+\gamma_{ge})\tau}\right] \tag{6.8}$$

and, using the steady-state value for $f_1(t)$ from Eq. 6.4, we get the result

$$g_2^e(\tau) = -e^{-(\gamma_{eg}+\gamma_{ge})\tau}, \tag{6.9}$$

where we use the superscript e to label this as the correlation function of the emitter's statistics alone. The full FCS correlation function is found by incorporating the systematics of the tracking-FCS curve described in Chapter 4 according to the multiplicative rule in Eq. 4.15. It is important to note that Φ cancels out of the correlation function; in effect, the sub-unity quantum yield only reduces the average fluorescence rate so that more photons must be detected to maintain a particular signal-to-noise ratio. It has no effect on the time-scales or functional forms of the functions $f_2(t, t+\tau)$ and $g_2(\tau)$.

The function $g_2^e(\tau)$ from Eq. 6.9 shows that the correlation between photons is reduced relative to the Poisson value (0) for time-scales $\tau < (\gamma_{eg} + \gamma_{ge})^{-1}$, reaching -1 at $\tau = 0$. This is a characteristic of anti-bunching statistics, indicating anti-correlation of photon detections at short times. $g_2^e(\tau)$ increases exponentially to 0, so that on long time-scales the statistics of the two-level emitter are indistinguishable from Poisson statistics. This tells us that our use of Poisson statistics in earlier chapters to describe the fluorescence statistics is not invalidated by anti-bunching because we were never concerned with such short time-scales. The long time-scale Poisson approximation is an example of *adiabatic elimination* of fast dynamics[82].

6.2.2 Computing waiting-time statistics

We now show that the result from Eq. 6.9 may also be computed using waiting time statistics, which is (in my opinion) a more elegant approach than simply evolving the master equation. We start by first finding the two-time joint probability density $p(t_k, t_{k+1})$ that gives the statistics of the waiting time between *adjacent* $\psi_e \to \psi_g$ transitions. This quantity differs from $f_2(t_k, t_{k+1})$ in that it specifies that no transitions occur between t_k and t_{k+1}, whereas $f_2(t_k, t_{k+1})$ allows any number of transitions to occur in that interval. $p(t_k, t_{k+1})$ is a true probability density; $f_2(t_k, t_{k+1})$ is not. We have

$$p(t_k, t_{k+1}) = p(t_{k+1}|t_k)p(t_k), \tag{6.10}$$

where $p(t_k)$ is just the steady-state detection probability from Eq. 6.4 (with $\Phi = 1$) and $p(t_{k+1}|t_k)$ represents the probability that, *given* that the emitter is in the ground state at time t_k, it jumps to its excited state at some time $\tau \in (t_k, t_{k+1})$ and then relaxes back to the ground state at time t_{k+1}. Since both of these transitions are defined by single, constant rates, they obey exponential statistics[8]. We use this fact to directly compute $p(t_{k+1}|t_k)$:

$$\begin{aligned} p(t_{k+1}|t_k) &= \int_0^{t_{k+1}-t_k} \mathrm{d}\tau \gamma_{eg} e^{-\gamma_{eg}\tau} \gamma_{ge} e^{-\gamma_{ge}(t_{k+1}-t_k-\tau)} \\ &= \frac{\gamma_{eg}\gamma_{ge}}{\gamma_{ge} - \gamma_{eg}} \left(e^{-\gamma_{eg}(t_{k+1}-t_k)} - e^{-\gamma_{ge}(t_{k+1}-t_k)} \right). \end{aligned} \tag{6.11}$$

Just to ensure the consistency of this result, we consider the apparent pole in the denominator of the constant factor for $\gamma_{eg} = \gamma_{ge}$:

$$\lim_{\gamma_{eg} \to \gamma_{ge}} = \gamma_{ge}^2 (t_{k+1} - t_k) e^{-\gamma_{ge}(t_{k+1} - t_k)}, \tag{6.12}$$

indicating that the pole is canceled by the zero in the term containing the exponentials.

It is a straightforward generalization of Eq. 6.11 to compute the conditional probability density of two transition times that are not necessarily adjacent. Let us compute the probability density that a transition occurs at time t_{k+n}, given a transition at time t_k, for $n > 0$. This is found by integrating over the intermediate times $t_{k+1}, \dots, t_{k+n-1}$ using the Chapman-Kolmogorov equation,

$$p(t_{k+n}|t_k) = \int_0^\infty \mathrm{d}t_{k+n-1} \cdots \int_0^\infty \mathrm{d}t_{k+1} p(t_{k+n}|t_{k+n-1}) \cdots p(t_{k+1}|t_k), \tag{6.13}$$

where the conditional probabilities are simplified because each detection time only depends on the detection time immediately prior to it. Equation 6.13 is a convolution integral that can be simplified by using the Laplace transform of Eq. 6.11,

$$\tilde{p}(s) = \frac{\gamma_{eg}\gamma_{ge}}{(s + \gamma_{eg})(s + \gamma_{ge})}, \tag{6.14}$$

and the convolution theorem. Equation 6.13 becomes

$$p(t_{k+n}|t_k) = \mathcal{L}^{-1}\left[\left(\frac{\gamma_{eg}\gamma_{ge}}{(s + \gamma_{eg})(s + \gamma_{ge})}\right)^n ; t_{k+n} - t_k\right]. \tag{6.15}$$

This inverse Laplace transform is not easy to calculate. Luckily, we will not need to invert it to compute the results that we require.

The function $f_2(t, t+\tau)\mathrm{d}t^2$ represents the probability that two $\psi_e \to \psi_g$ transitions occur in the intervals $(t, t + \mathrm{d}t)$ and $(t + \tau, t + \tau + \mathrm{d}t)$ with *any number* of transitions occurring at times between t and $t + \tau$, and that both of these transitions are radiant. We set $t_k = t$ and then compute $f_2(t, t+\tau)$ by summing the probabilities that $t + \tau = t_{k+1}$, $t + \tau = t_{k+2}$, and so on. We have

$$\begin{aligned} f_2(t, t+\tau) &= \Phi^2 \sum_{n=1}^{\infty} p(t_{k+n} = t + \tau | t_k = t) p(t_k = t) \\ &= \frac{\Phi^2 \gamma_{eg}\gamma_{ge}}{\gamma_{eg} + \gamma_{ge}} \mathcal{L}^{-1}\left[\sum_{n=1}^{\infty}\left(\frac{\gamma_{eg}\gamma_{ge}}{(s + \gamma_{eg})(s + \gamma_{ge})}\right)^n ; \tau\right], \end{aligned} \tag{6.16}$$

where we have inserted Eq. 6.15 and brought the sum inside the Laplace transform. The series converges because the inverse transform is taken along a contour that is fully contained in the right half of the complex plane (so that it is the geometric series), so we have

$$\begin{aligned} f_2(t, t+\tau) &= \frac{\Phi^2 \gamma_{eg}\gamma_{ge}}{\gamma_{eg}+\gamma_{ge}} \mathcal{L}^{-1}\left[\frac{\gamma_{eg}\gamma_{ge}}{s(s+\gamma_{eg}+\gamma_{ge})}; \tau\right] \\ &= \left(\frac{\Phi\gamma_{eg}\gamma_{ge}}{\gamma_{eg}+\gamma_{ge}}\right)^2 \left[1 - e^{-(\gamma_{eg}+\gamma_{ge})\tau}\right]. \end{aligned} \tag{6.17}$$

Clearly, since this is the same as Eq. 6.8, it gives rise to the same correlation function. This approach required more thought about the statistics of emission than the master equation evolution did, and rewarded us by not requiring a messy calculation of a matrix exponential.

6.2.3 Effect of background counts

In any real experiment we will inevitably detect background photons, the statistics of which are usually Poissonian and completely uncorrelated with those of the emitter (or at least they are very loosely correlated, since small fluctuations in the laser intensity affect the statistics of both background scattering and fluorescence). Because of this background the probability of coincident photon detections does not actually go to zero as $\tau \to 0$, since two background photons can be detected simultaneously, or one photon from background and one from the emitter can also be detected simultaneously. We now compute the quantitative influence of this effect on the correlation function.

Just as in Section 4.1.4, we assume that the background is a Poisson process with constant rate Γ_B. We use Eq. 4.17 to compute the FCS curve, inserting $f_2(t, t+\tau)$ from Eq. 6.17 and $f_1(t)$ from Eq. 6.4. This gives us

$$g_2^e(\tau) = -\left[\frac{\gamma_{eg}\gamma_{ge}}{\gamma_{eg}\gamma_{ge} + \Gamma_B\left(\gamma_{eg}+\gamma_{ge}\right)}\right]^2 e^{-(\gamma_{eg}+\gamma_{ge})\tau}. \tag{6.18}$$

Equation 6.18 is more clearly expressed in terms of the ratio of the background count rate to the emission rate. If we define

$$\kappa \equiv \Gamma_B \left(\frac{\gamma_{eg}+\gamma_{ge}}{\gamma_{eg}\gamma_{ge}}\right), \tag{6.19}$$

then Eq. 6.18 becomes

$$g_2^e(\tau) = -\left(\frac{1}{1+\kappa}\right)^2 e^{-(\gamma_{eg}+\gamma_{ge})\tau}. \tag{6.20}$$

This result tells us that background noise will prevent the correlation function from reaching its minimum value -1 at $\tau = 0$.

6.2.4 Effect of additional emitters

The presence of a second emitter will have a dramatic effect on the depth of the anti-bunching dip in the correlation function because the second emitter is just as bright as the first. Equation 6.20 predicts that for a Poissonian emitter with equal brightness as the initial emitter ($\kappa = 1$), the anti-bunching dip will be only 25% as deep as for a single emitter. However, if the second emitter is a two-level system then it too will exhibit anti-bunching, so the effect will be less pronounced than this.

Here we use the general expression Eq. 4.16 and insert identical values for the f_n from both the emitter and the background. The resulting correlation function is

$$g_2^e(\tau) = -\frac{1}{2}\, e^{-(\gamma_{eg}+\gamma_{ge})\tau}, \tag{6.21}$$

which shows that the anti-bunching dip becomes half as deep in the presence of a second two-level emitter — a reduction that is half as large as it would be if the second emitter were Poissonian.

It is easy to generalize to N emitters, and even further to N emitters that are not all equally bright. For notational simplicity we denote the average fluorescence rate of emitter j by b_j, and the anti-bunching recovery rate $\gamma_{ge} + \gamma_{eg}$ for emitter j by γ_j. We assume a constant Poisson background Γ_B. It is straightforward to show that the correlation function is given by

$$g_2^e(\tau) = -\frac{\sum_{j=1}^{N} b_j^2 e^{-\gamma_j \tau}}{\left(\sum_{j=1}^{N} b_j\right)^2 (1+\kappa)^2}, \tag{6.22}$$

where, as in Eq. 6.19, κ is the ratio of the background count rate to the fluorescence count rate,

$$\kappa = \Gamma_B \left(\sum_{j=1}^{N} b_j \right)^{-1}. \tag{6.23}$$

This correlation function will be useful to us when we study the excitation of multiple electrons within a single qdot in Section 6.3.3.

6.2.5 Implications of the Cauchy-Schwarz inequality: photons!

In deriving anti-bunching statistics, we assumed that the fluorescence light is radiated in delta-function bursts, not in packets of any finite duration. This assumption is taken for granted in modern physics, because no one denies the existence of photons. However, the resulting correlation function, with $g_2^e(0) = 0$, *cannot* be generated by bursts of light with any nonzero correlation time. This means that the observation of photon anti-bunching is experimental proof that light exists as quantized photons, not a continuous electromagnetic wave. Moreover, it can be shown that the oft-cited "proof" via the photoelectric effect — for which Einstein was awarded his Nobel prize — is actually insufficient to prove photons' existence because quantized electron energy and classical radiation would also produce the effect. We do not discuss this here, but the argument is expounded in [80].

Instead of delta-function bursts, we can assume that the emitter radiates bursts that have a finite correlation time or, more generally, have a pulse shape $h(t)$. We can write the magnitude of the fluorescence emission during an experiment of duration T as

$$I(t) = \sum_{n=1}^{N} h(t - t_n), \tag{6.24}$$

where $0 \leq t_n \leq T$ is the list of times at which the emitter relaxed to the ground state and emitted a fluorescence burst. Since the radiated energy is both finite and proportional to the square of the magnitude, we know that $h(t)$ must be square integrable:

$$\int_{-\infty}^{\infty} h(t)^2 \, \mathrm{d}t < \infty. \tag{6.25}$$

Then $h(t)$, and consequently $I(t)$, lie in the L^2 inner product space over the real line[83, e.g.]. The Cauchy-Schwarz inequality [83] states that, for any x and y in an inner product space,

$$\langle x, y \rangle \leq \|x\| \|y\|, \tag{6.26}$$

so in our case we have

$$\begin{aligned} \int_{-\infty}^{\infty} I(t) I(t+\tau) \, \mathrm{d}t &\leq \left(\int_{-\infty}^{\infty} I(t)^2 \, \mathrm{d}t \right)^{1/2} \left(\int_{-\infty}^{\infty} I(t+\tau)^2 \, \mathrm{d}t \right)^{1/2} \\ &= \int_{-\infty}^{\infty} I(t)^2 \mathrm{d}t. \end{aligned} \tag{6.27}$$

The fluorescence signal is causal ($I(t) = 0$ for $t < 0$ and $t > T$) — we do not detect any fluorescence when our detector is not on. In terms of the unnormalized correlation function

$$G_2(\tau) = \langle f_2(t, t+\tau) \rangle \,, \tag{6.28}$$

where the averages are taken over time, Eq. 6.27 is written

$$G_2(\tau) = \int_0^T I(t) I(t+\tau) \, \mathrm{d}t \leq \int_0^T I(t)^2 \, \mathrm{d}t = G_2(0). \tag{6.29}$$

This inequality tells us that our assumption — that the radiated fluorescence is square-integrable — must be false if we observe anti-bunching in any real system. However, if that radiation were a classical electromagnetic wave, this would imply that each burst transmitted an infinite amount of energy. This is unacceptable, leading us to the conclusion that the radiation must *not* be a classical electromagnetic wave, but rather a quantized sequence of bursts of infinitesimal width.

It is interesting to consider what the correlation function would look like if $h(t) \neq \delta(t)$, especially because every electronic signal that we measure in the lab is classical (we cannot detect a signal from a single electron because such a signal is dwarfed by electronic noise). We have

$$G_2(\tau) = \left\langle \frac{1}{T} \int_0^T \sum_{i,j} h(t-t_i) h(t+\tau-t_j) \mathrm{d}t \right\rangle , \tag{6.30}$$

which in general is not very easy to calculate, even for simple $h(t)$, because we must account for the overlap of all possible numbers of pulses from adjacent photon detections. However, things are simplified if the pulse width w is very short compared to the time spacing $(\gamma_{eg} + \gamma_{ge})(\gamma_{eg}\gamma_{ge})^{-1}$ between detected pulses, so that the probability of two pulses overlapping is very small. If we assume, for simplicity, that the pulses are top hat-shaped and their integral is one, then the normalized FCS curve is

$$g_2^e(\tau) \approx 1 - e^{-(\gamma_{eg}+\gamma_{ge})|\tau|} + \Theta\left(w - |\tau|\right) \frac{\gamma_{eg} + \gamma_{ge}}{\gamma_{eg}\gamma_{ge}} \frac{1}{w^2} \left(w - |\tau|\right) , \tag{6.31}$$

where Θ is the Heaviside step function. Due to our assumption of short pulses, $g_2^e(0) \gg 1$ and the Cauchy-Schwarz inequality is satisfied. This tells us that the self-correlation of $h(t)$ dominates $g_2^e(\tau)$ on short time-scales, so that no sequence of classical pulses can exhibit sub-Poissonian statistics when directly correlated with itself. We are able to detect anti-bunching in our experiment because we *interpret* the signal we detect as coming from a photon-emitting source, where the nonzero pulse width is an artifact of the acquisition electronics. Figure 6.2 shows the difference between the correlation functions for photon anti-bunching and for anti-bunching of pulses with finite width.

6.3 Experimental measurements of quantum dot anti-bunching

We purchased carboxy-derivatized qdots with 655nm emission peaks from Invitrogen/Molecular Probes (Qdot655; Eugene, OR). We prepared a buffer consisting of 50mM sodium borate in deionized water and adjusted the pH to 8.3, as measured on a Beckman Φ45 meter with a saturated KCl probe, by titration with sodium hydroxide. Immediately before adding qdots to the buffer we added 140mM 2-mercaptoethanol in order to suppress qdot blinking[76]. We

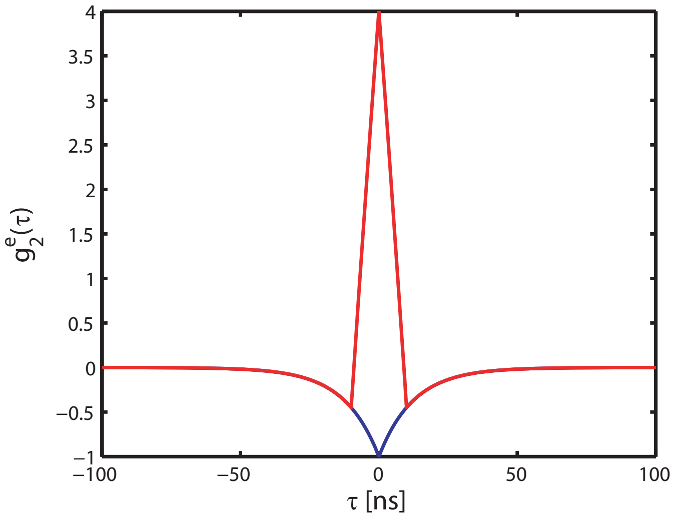


Figure 6.2: $g_2^e(\tau)$ for anti-bunched photons (blue) and pulses with w =10ns (red). Both curves used $\gamma_{eg} = \gamma_{ge} = 4 \cdot 10^7 \mathrm{s}^{-1}$.

diluted the qdots to 80fM in this buffer; this extremely low concentration ensured that qdots drifted into focus of the apparatus very rarely, so that our measurements on each qdot were not contaminated by the detection of a second qdot.

6.3.1 Data

Figure 6.3 shows the raw data from a sample tracking run. The qdot is tracked for over 20 seconds, after which it is lost because the z tracking stage reached its upper boundary and all stages were reset. The fluorescence remains roughly constant throughout the run, but several large dips are visible (the most dramatic of which are at $t \approx 7$s and $t \approx 19$s). In order to determine whether to attribute the dips in fluorescence to tracking error or qdot blinking, we computed the mean-squared deviations of the tracking stage positions (see Sections 3.2.2 and 5.3.5 and [42]) and fit curves to this data. The results gave us an estimate of the qdot's diffusion coefficient $D \approx (19.5 \pm 2.0)\mu\ \mathrm{m}^2/s$, where the uncertainty comes from averaging over all three axes. This diffusion coefficient corresponds roughly to a 20nm sphere. In addition, we find the RMS tracking error to be about 410nm along the x axis, 260nm along the y axis and 280nm along the z axis. The relatively large error on the x axis was the result of a slight asymmetry in the beam rotation that caused a larger slope of the localization signal along that axis. This caused the loop gain of the tracking system to be greatest on that axis, and resulted in excess noise appearing in the stage positions. Our tracking beams had waists of about 1.2μm and 1.4μm, implying that even a 2σ tracking error in the radial direction would result in a drop in fluorescence intensity of about 50%. It is very unlikely that these nearly 100% drops in fluorescence intensity are due to tracking errors. We attribute them to qdot blinking, which we

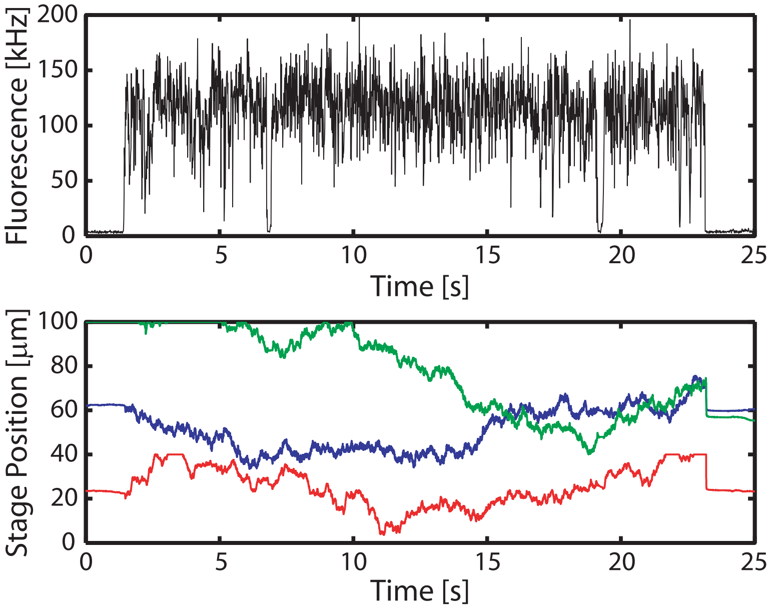


Figure 6.3: Qdot tracking in 50mM sodium borate/140mM 2-mercaptoethanol (and no glycerol). Top: Fluorescence intensity, in 10ms bins. Bottom: Positions of the x (blue), y (green) and z (red) tracking stages. The y position is initially beyond the range of the data-acquisition device.

study in more depth in later sections.

Given the diffusion coefficient that we measured and the size of our laser beams, the open-loop FCS diffusion time of these qdots through the focus of our laser was about $\tau_D \approx 30$ms. We tracked qdots for roughly three orders of magnitude longer than this, limited almost entirely by the travel range of the tracking stages. In open-loop FCS the average fluorescence rate of a particle in the effective sampling volume $\bar{V}$, as defined in Eq. 4.13, is only about 35% of the peak fluorescence rate at the origin. In contrast, the average fluorescence rate of a qdot that we tracked with the tracking error reported above is greater than 80% of the peak rate. The resulting $g_2(\tau)$ curves have between fifty-fold and hundred-fold improvements in their signal-to-noise ratio, and this was sufficient for measuring anti-bunching at mean fluorescence rates as low as $65kHz$.

We tracked 80 quantum dots, for between 3.5s and 29s, over a range of excitation intensities that was controlled by attenuating the excitation beam with neutral density filters. We computed the FCS curve for each of these qdots for $|\tau| < 200$ns and fit the result to the function

$$g_2(\tau) = \{1 - \alpha e^{-\gamma\tau}\}\, g_2^0(0), \tag{6.32}$$

where α accounts for attenuation of the anti-bunching dip due to background counts or multiple electron excitations[84]. $g_2^0(0)$ represents the contributions to the curve other than anti-bunching, all of which have much longer characteristic time-scales and so can be safely treated

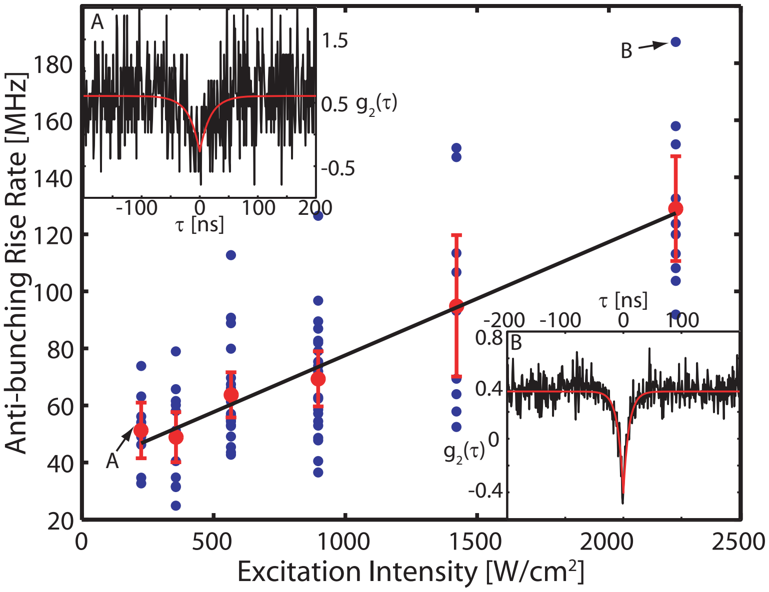


Figure 6.4: Rise rates $\gamma \equiv \gamma_{eg} + \gamma_{ge}$ (blue circles) as determined from 80 tracked qdots at varied excitation intensities. Red circles with error bars indicate the sample mean. Insets show $g_2(\tau)$ for data points A and B with fits to Eq. 6.32. Linear fit: $\gamma = (37.8 \pm 8.8) + (3.97 \pm 0.8) \cdot 10^{-2} I$ [MHz].

as a constant over the time-scale of these fits. In Fig. 6.4 we show the anti-bunching rise rates $\gamma \equiv \gamma_{eg} + \gamma_{ge}$ extracted from the fits, in addition to the anti-bunching curves for two example qdots at different excitation intensities. The excitation intensity was inferred from the optical power in each beam and the irradiance profiles of the beams as measured on an immobilized fluorescent bead. We estimated a 15% loss in power through the objective, immersion water and coverslide.

We expect γ to vary linearly with excitation intensity because the excitation rate γ_{eg}, given in Eq. 6.1, is proportional to the intensity. A linear fit to the data is shown in Fig. 6.4 and predicts an average fluorescence lifetime $\tau_f \equiv \gamma_{ge}^{-1} = (27 \pm 6)$ns. This value is comparable to the average lifetimes 20ns[77] and (32 ± 2)ns[84] measured on surface-immobilized CdSe/ZnS qdots; we do not see a statistically-significant deviation in our data from previous measurements, although the difference between those measurements implies that deviation between samples may be even larger than our experimental uncertainty.

The slope of the linear fit gives $\sigma_a = (1.48 \pm 0.15) \cdot 10^{-14}\text{cm}^2$. From this σ_a we may infer the bulk extinction coefficient ϵ of a solution of qdots in the following way. In a bulk measurement, a beam of cross-section A traverses a length L of sample, and we measure the fraction of the incident power that is absorbed. ϵ is the ratio of the absorbed to the incident power, per unit L, for a sample concentration C of 1M. In the volume AL there are $N_A CAL$ qdots, where N_A is the Avogadro number. These qdots absorb the light incident upon the area $\sigma_a N_A CAL$, so that

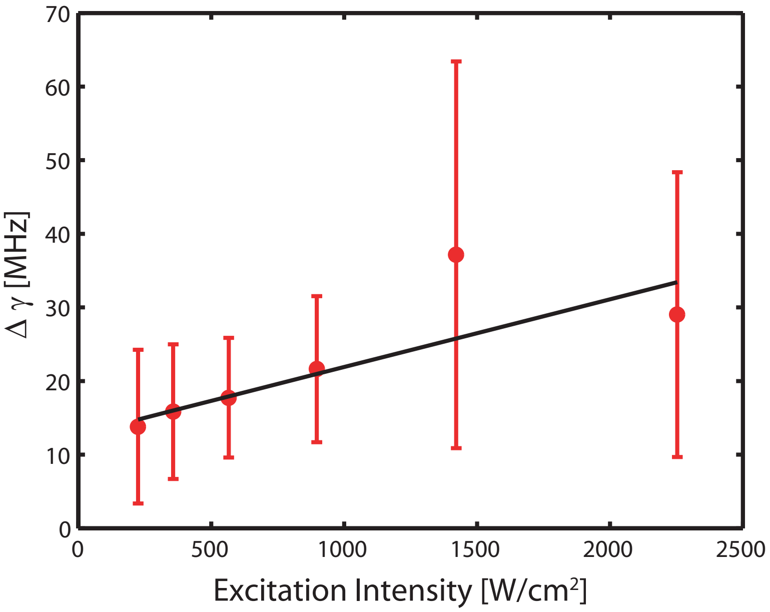


Figure 6.5: Standard deviation $\Delta\gamma$ of the anti-bunching recovery rates at different values of the excitation intensity I. 2σ error bars were computed using Eq. 6.33. Linear fit (by maximum-likelihood method): $\Delta\gamma = (12.7 \pm 5.8) + (9.2 \pm 7.0) \cdot 10^{-3} I$[MHz].

the fraction absorbed per unit L is $\sigma_a N_A C$. Inserting a concentration of 1M and converting the units, the σ_a we measured implies a bulk extinction coefficient $\epsilon = (8.9 \pm 0.9) \cdot 10^6 \mathrm{M}^{-1}\mathrm{cm}^{-1}$. This is significantly larger than the the value of $2.1 \cdot 10^6 \mathrm{M}^{-1}\mathrm{cm}^{-1}$ specified by the manufacturer. It is possible that the enhancement is due to the addition of 2-mercaptoethanol to our qdot solutions, but we cannot be certain that this was the cause.

6.3.2 Recovery rate heterogeneity

Perhaps the most distinguishing feature of the data in Fig. 6.4 is the large variation in γ observed for different qdots at the same excitation intensity. We computed the relative standard deviation $\Delta\gamma/\gamma$ of the rise rate and found that it varies little with intensity, ranging between 23% and 40% with a mean of (30 ± 5)%. This implies that the spread is not due to poor signal, since the signal-to-noise ratio improves as the intensity increses. It is also not due to intensity fluctuations, as the power in our beam is stable to within 5% over time-scales longer than any of our data runs. We conclude that the variation we see is primarily due to heterogeneity in the qdot sample. Such heterogeneity has been measured before on qdots immobilized in glass[77], but it was impossible to determine from those measurements whether this effect was an inherent property of the qdot sample or the result of their proximity to notoriously reactive glass surfaces[81]. Our results provide strong evidence that the heterogeneity *is* a property of the qdot sample, not an experimental artifact.

To further characterize the heterogeneity of the qdots, we can try to determine the relative contributions of variations in γ_{ge} and σ_a from our data by examining the intensity dependence of the variations in γ. In Fig. 6.5 we plot $\Delta\gamma$ as a function of I from the data in Fig. 6.4. Each data point in the plot is an *estimate* of the sample standard deviation using the 10-20 data points at each excitation intensity. Like all statistical estimates[85], the standard deviation estimate has an intrinsic variance that depends on the sample size. If we assume that the distribution of γ about its mean at any fixed intensity is Gaussian with variance $(\Delta\gamma)^2$, then the variance of the standard deviation estimator is

$$\mathrm{Var}\left\{\left[\frac{1}{N-1}\sum_{j=1}^{N}\left(\gamma_j-\langle\gamma\rangle\right)^2\right]^{1/2}\right\}=\frac{(\Delta\gamma)^2}{N-1}. \tag{6.33}$$

We use this variance to compute the 2σ error bounds shown in Fig. 6.5.

We expect $\Delta\gamma$ and I to be linearly related, so we fit a line to the data in Fig. 6.5 in order to extract the standard deviations $\Delta\gamma_{ge}$ and $\Delta\sigma_a$ from

$$\Delta\gamma=\Delta\gamma_{ge}+10^{-6}\frac{\lambda}{hc}I\Delta\sigma_a. \tag{6.34}$$

Due to the large uncertainty in each data point an ordinary least-squares fit would be inappropriate; our fit needs to account for this uncertainty. We computed the most probable linear fit to the data using a numerical maximum-likelihood method. Figure 6.6 shows the likelihood function

$$L\left(\Delta\gamma_{ge},\Delta\sigma_a\right)=\prod_j\exp\left[-\frac{\left(N_j-1\right)\left(\Delta\gamma_j-\Delta\gamma_{ge}-\frac{10^{-6}\lambda}{hc}I_j\Delta\sigma_a\right)^2}{2\left(\Delta\gamma_j\right)^2}\right], \tag{6.35}$$

where $(I_j,\Delta\gamma_j)$ are the six data points in Fig. 6.5 and N_j is the number of data points at each intensity value in Fig. 6.4. The parameters defining the line in Fig. 6.5 were those that maximized the likelihood function, and the uncertainties were computed by finding the rectangular box centered at the maximum that contained 95% of the likelihood. Due to the skew orientation of the elliptical contours, these uncertainties are somewhat conservative: a more careful approach would characterize the $\Delta\gamma_{ge}$ and $\Delta\sigma_a$ as lying inside the skew ellipse containing 95% of the likelihood. Such a characterization, however, would not give independent bounds for the two parameters.

The maximum-likelihood fit gave $\Delta\gamma_{ge}=(12.7\pm5.8)$MHz. This is a reasonably tight uncertainty bound, and indicates that the qdots that we studied almost certainly exhibited variations in γ_{ge} because zero is far outside of this range. The corresponding standard deviation in the fluorescence lifetime, using $\gamma_{ge}=37.8$MHz from the fit above, is $\Delta\tau_f=(8.9\pm4.1)$ns. This number compares favorably with the lifetime variation measured on surface-immobilized qdots

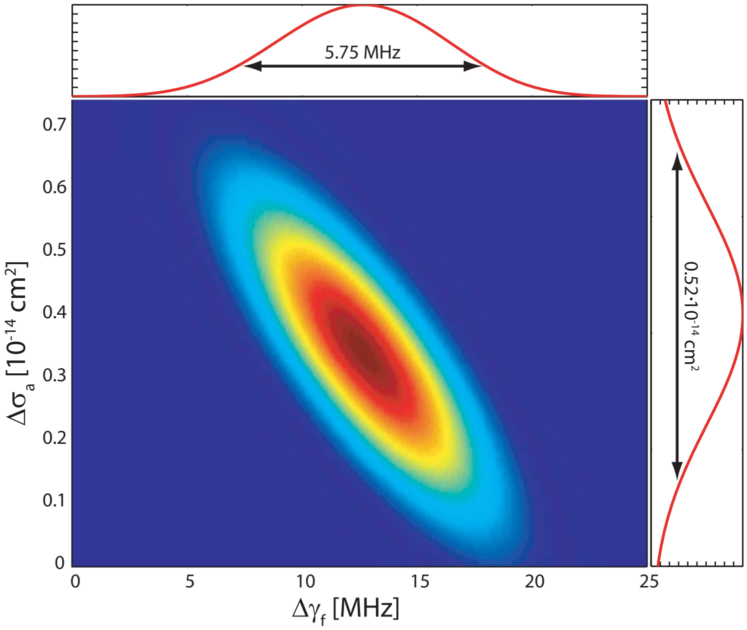


Figure 6.6: Likelihood function from Eq. 6.35 for the standard deviations $\Delta\gamma_{ge}$ and $\Delta\sigma_a$. The peak at $\Delta\gamma_{ge} = 12.7$MHz and $\Delta\sigma_a = 0.34 \cdot 10^{-14}\text{cm}^2$ represents the linear fit shown in Fig. 6.5. The plots above and to the right of the contour plot are marginalized likelihood functions found by integrating over $\Delta\sigma_a$ and $\Delta\gamma_{ge}$, respectively. Uncertainties in the linear fit parameters are found by computing the 2σ widths of the marginal likelihood functions.

($\Delta\tau_f \approx 7$ns)[77], so that we have no evidence that the size of the variation in those measurements can be attributed to interactions between the qdot and the glass surface.

The fit $\Delta\sigma_a = (0.34 \pm 0.26) \cdot 10^{-14}\text{cm}^2$ was not as convincing: while zero is outside of the 2σ bound, it is well within 3σ and so there is a small ($\sim$ 2%) probability that σ_a is constant among the different qdots in the sample. The size of $\Delta\sigma_a$ is less than about 40% of the value of σ_a deduced from the data in Fig. 6.4. This relative variation is considerably smaller than that measured on glass-immobilized qdots (~80%)[77]. The qdots used in that study clearly differed from ours in their absorptive properties, as their bulk extinction coefficient was almost two orders of magnitude smaller than ours at an even lower excitation wavelength. While we can certainly say that we observe quantitative differences from those measurements, we cannot safely attribute the differences to the the presence of the glass. In future work we will examine the differences between the photon statistics of qdots from identical samples both free in solution and on glass surfaces with a variety of surface chemistries.

In order to be able to say with greater confidence whether the qdots we studied exhibit variations in σ_a at all (to rule out the possibility that $\sigma_a = 0$), we require more data. Even a fairly modest 4-fold increase in the number of qdots tracked at each excitation intensity would likely reduce the uncertainty in $\Delta\sigma_a$ to a level that would allow us to answer this question. Recording fluorescence from qdots at a wider range of intensities than measured here would help as well.

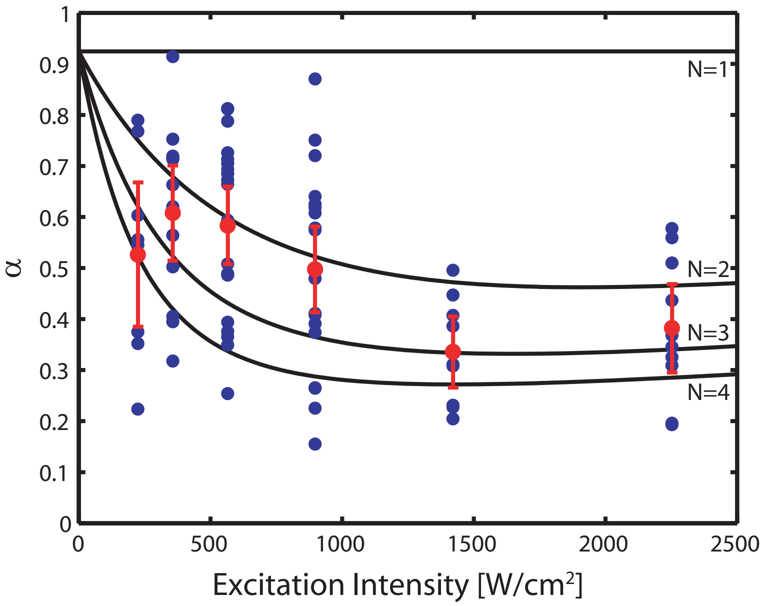


Figure 6.7: The attenuation factor α from Eq. 6.32 plotted for each of the $g_2(\tau)$ fits in Fig. 6.4. Red circles indicate the mean α at each intensity, with 2σ error bars about the means. Black curves indicate predicted values for a qdot with 1, 2, 3, and 4 excitable electrons according to the model described in the text.

Alternatively, the wide range in tracking trajectory durations for which we were able to resolve anti-bunching indicates that we often acquire far more data than needed for this measurement. We could modify our apparatus to change the excitation intensity while tracking a single qdot once we have acquired enough data. For long tracking trajectories we could essentially map out the intensity dependence of γ for each qdot, making it easy to determine whether there is variation in σ_a between qdots or not.

6.3.3 Evidence of multiple excitations

The $g_2(\tau)$ curves in Fig. 6.4 contain anti-bunching dips at $\tau = 0$ that are shallower than predicted by the theory in Section 6.2 for a single emitter in the absence of background. Figure 6.7 shows the fitted values of α for all of the qdots that we tracked. It is clear from the figure that the shallow anti-bunching dips in Fig. 6.4 were not an anomaly; there is quite a bit of variation in the fitted values for α, which range from between about 0.15 and over 0.9.

In Section 6.2.3 we showed how background counts will always ensure that $\alpha < 1$. To account for background in our data we computed the average ratio of the size of the background to the size of the fluorescence from each qdot, and found $\kappa = (4 \pm 0.6) \cdot 10^{-2}$ when averaged over all qdots. According to Eq. 6.20, this κ imposes an upper limit of 0.92 ± 0.01 on the attainable value for α, corresponding to the line at $N = 1$ in Fig. 6.7. α nearly reached this maximum for two qdots but the remaining fits all produced substantially lower α values, near

0.6 for excitation intensities less than 1kW/cm^2 and 0.4 for higher intensities.

We used a very simple model to attempt to explain the deviation in α from the optimal theoretical value as well as the apparent intensity-dependence of the α values. We allow up to a finite number N of electrons to be excited simultaneously in the qdot. Each electron is excited at the same rate γ_{eg} and relaxes at the same rate γ_{ge}. We let ψ_j denote the state in which the qdot contains j excited electrons. The allowable transitions between states ψ_j are illustrated in Fig. 6.8. The $\psi_j \to \psi_{j-1}$ transition occurs with rate $j\gamma_{ge}$ because any of the j excited electrons may relax to produce the transition. Similarly, the upward transition $\psi_j \to \psi_{j+1}$ occurs at the rate $(N-j)\gamma_{eg}$ because there are $N-j$ ground-state electrons available to be excited. It can be shown that the steady-state population ρ_j^∞ of state ψ_j is given by

$$\rho_j^\infty \propto \begin{pmatrix} N \\ j \end{pmatrix} \gamma_{ge}^{N-j}\gamma_{eg}^{j}, \tag{6.36}$$

where we used the standard notation for the binomial coefficient. We may then express the average fluorescence rate of the $\psi_j \to \psi_{j-1}$ transition by

$$b_j = j\gamma_{ge}\rho_j^\infty \propto j \begin{pmatrix} N \\ j \end{pmatrix} \gamma_{ge}^{N-j+1}\gamma_{eg}^{j}. \tag{6.37}$$

These b_j were first discussed in Section 6.2.4 when we derived the contribution of additional emitters to the anti-bunching correlation function. Equation 6.37 may appear misleading because the proportionality constant contains terms that depend on γ_{eg} and will cause the fluorescence rate to saturate; however, those terms disappear from the correlation function in Eq. 6.22 so they do not contribute to α. We use Eq. 6.22 to compute the predicted value of α for $N = 1$, 2, 3, and 4 and plot them in Fig. 6.7.

The curves that we derived from this model seem to capture the qualitative features of the data. Their rapid decay at low excitation intensities explains why we rarely see α approaching its upper limit even at low excitation intensities. The slight increase in the theory curves at higher excitation intensities occurs when $\gamma_{eg} > \gamma_{ge}$, so that upward transitions dominate downward transitions. As γ_{eg} increases, $\psi_N \to \psi_{N-1}$ eventually begins to dominate all other transitions, resulting in an enhancement in the anti-bunching dip.

The curve for $N = 2$ fits the data well at excitation intensities below 1kW/cm^2, but overpredicts the data at higher intensities. The curves at $N = 3$ and $N = 4$ match the data at low (<0.5kW/cm^2) and high (>1kW/cm^2) intensities, but underpredict the data at intermediate intensities. The curves suggest that more data should be taken at lower excitation intensities to try to observe the predicted strong dependence of α on I. This may be difficult, however, due to

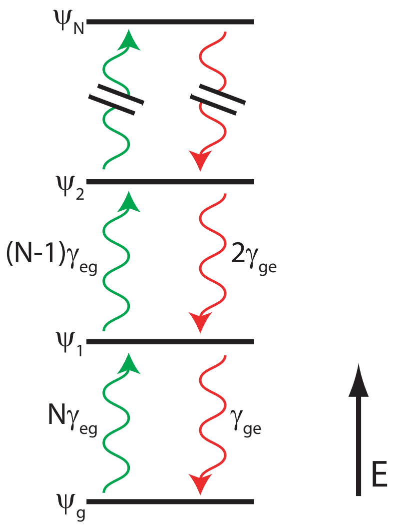


Figure 6.8: Simple multiple-electron qdot model. We let ψ_j be the state of the qdot with j excited-state electrons and $N-j$ ground-state electrons. Each electron transition is independent of each other and occurs with the same upward and downward rates, so that the upward and downward rates for the qdot state transitions are equal to the single electron transition rates scaled by a combinatorial factor as described in the text.

the large numbers of photons needed to develop good statistics on anti-bunching time-scales.

The theory curves in Fig. 6.7 support the attribution of the reduced values of α to multiple electronic excitations in the qdot. However, our multiple-excitation model is oversimplified. To say any more about what we have measured we require a more appropriate model accounting for the complicated physics of multiply-excited qdots[86], and we probably require more data as well.

6.3.4 Absolute detection efficiency calibration

As an interesting application of our measurements in this section, we can estimate the absolute collection efficiency of our apparatus. Our measurement of the absorption cross-section σ_a allows us to determine the rate at which the qdot cycled between its ground and excited states. This is related to the rate at which the qdot was actually *emitting* photons by the fluorescence quantum yield; in the absence of blinking the quantum yield for CdSe/ZnS qdots is nearly unity, so we may estimate the emission rate very closely.

The fluorescence rate that we *collect* is related to the emission rate by a constant factor η that contains all of the systematic losses in the experiment. Figure 6.9 illustrates what is probably the largest of these. The fluorescence radiation is spherically symmetric, and the microscope objective only collects light from a second-order cone of angle θ, determined by its numerical aperture. The fraction of the fluorescence light that we collect is given by the ratio

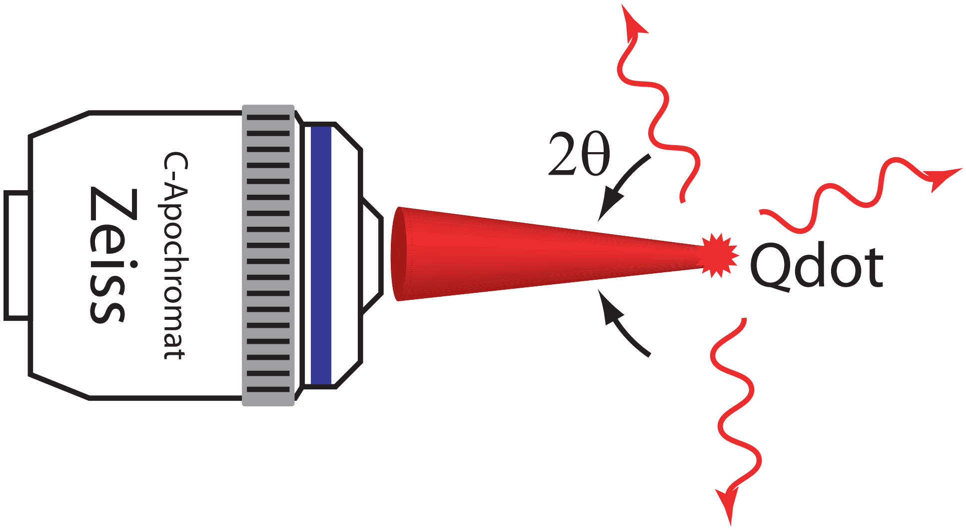


Figure 6.9: Photon collection from a single qdot. Fluorescence intensity is spherically-symmetric. The apparatus only collects light radiated within the cone in the drawing. The cone angle θ depends on the numerical aperture of the collection optics according to Eq. 6.38.

of the area subtended by the cone to the total surface area of the sphere:

$$\frac{A_{cone}}{A_{sphere}} = \frac{1}{2}\left(1-\cos\theta\right) \tag{6.38}$$

$$= \frac{1}{2}\left[1-\left(1-\frac{\mathrm{NA}^2}{n^2}\right)^{1/2}\right], \tag{6.39}$$

where n is the index of refraction of the medium (1.33 for water) and NA= $n\sin\theta$ is the numerical aperture of the objective. Even with our high numerical aperture water immersion objective, this limits us to a collection efficiency of around 28%. Additional losses occur all over the optical apparatus: scattering of fluorescence by the coverslides and immersion water accounts for a few percent; transmission losses in the microscope objective are about 20%; the best chromatic filters lose about 10% each; and photon-counting APDs have a detection efficiency of only about 65% at 655nm. The largest and most unfortunate loss is due to the mistaken use of the CVI Laser harmonic separator (see Section 5.2.3) in the fluorescence path when the qdot measurements were made. Based on its specifications, we expect fluorescence collection losses of 80-90% at 655nm.

Figure 6.10 shows the fluorescence rates of each tracked qdot, averaged over the entire duration of each tracking trajectory. There is a clear dependence on intensity that appears roughly linear at low intensities and saturates at higher intensities. The fluorescence rate we

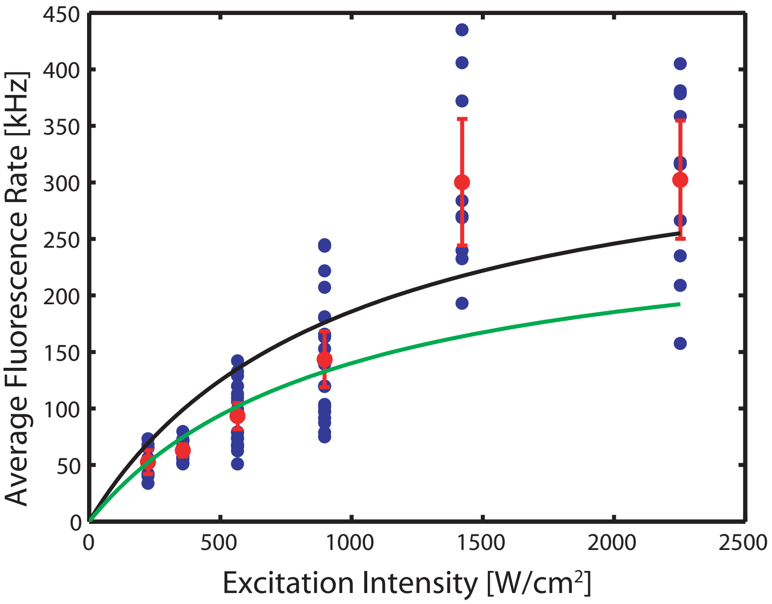


Figure 6.10: Qdot fluorescence rates averaged over the entire duration of each tracking trajectory. Solid curves show fits to the *detected* fluorescence rate as a fraction of the total fluorescence rate derived from Eq. 6.40, using the values γ_{ge} and γ_{eg} extracted from the data in Fig. 6.4. Fits to all of the data (black) yielded $\eta\Phi = 9.6 \cdot 10^{-3}$ (black) and fits to the data at the first four excitation intensities yielded $\eta\Phi = 7.2 \cdot 10^{-3}$ (green).

expect to collect, given the fits for γ_{ge} and σ_a, is computed using Eq. 6.4. This gives

$$f_1(t) = \Phi\eta \left(\frac{1}{\gamma_{ge}} + \frac{10^6 hc}{\sigma_a \lambda I} \right)^{-1}, \tag{6.40}$$

so that the product $\Phi\eta$ is the only free parameter. We fit Eq. 6.40 to the data, yielding $\Phi\eta = 0.96\%$, and the result is the black curve shown in the figure. The fit curve matches the data rather unsatisfactorily: it is too large at low excitation intensities and too small at high intensities. This may be due to a nonlinearity in the fluorescence rate; as discussed in Section 6.3.3, we see evidence that we excited more than one electron within the qdots. At higher excitation intensities more electrons are excited, which leads to a nonlinear dependence of the fluorescence rate on the excitation intensity. We do not attempt to account for this here; instead, we excluded the data from the two highest excitation intensites and computed another fit, giving $\Phi\eta = 0.72\%$. This curve is plotted in green; it fits the data at low intensities very well, but underpredicts the fluorescence rate at the high intensities.

If we assume a fluorescence quantum yield near unity, the fitted values of $\Phi\eta$ indicate that we collect approximately 1% of the fluorescence emitted by the qdot. We estimate that our collection optics and chromatic filters account for a loss of about 80-85%, and the detector quantum efficiency brings this number up to about 90%. The harmonic separator's transmissive losses further increase this to 98-99%, so that our 1% estimate is not at all unreasonabe for the apparatus as it was used to collect the qdot data.

6.4 Blinking statistics

Blinking — the turning off and on of the fluorescence signal at random intervals — is a hallmark of quantum dot fluorescence and probably the most important deviation of qdots from ideal fluorescence emitters[71]. In this process, an excited-state electron makes a nonradiative transition to a "trap" state at the surface of the qdot, effectively ionizing the qdot. Subsequent excitations of other electrons in the qdot decay rapidly via a nonradiative Auger process[87] in which their energy is transferred to the trapped electron rather than to an emitted photon. As a result the quantum yield of the ionized qdot is essentially zero, so its fluorescence disappears. Eventually the trapped electron is released back to the excited state, and the qdot's fluorescence returns.

6.4.1 Three-state model

A blinking qdot can be described very simply by introducing a third, dark state to the two-level emitter model[72]. Figure 6.11 illustrates this. Transitions between the excited state ψ_e and the dark state ψ_0 cause the qdot to blink on and off. If we let $\rho = \begin{pmatrix} \rho_g & \rho_e & \rho_0 \end{pmatrix}^T$ be the occupancy probabilities of each state, then the states evolve according to

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho = \begin{pmatrix} -\gamma_{eg} & \gamma_{ge} & 0 \\ \gamma_{eg} & -(\gamma_{ge} + \gamma_{0e}) & \gamma_{e0} \\ 0 & \gamma_{0e} & -\gamma_{e0} \end{pmatrix} \rho. \tag{6.41}$$

The three-state model can be simplified a bit because the fluorescence lifetime is much shorter than the average blinking time ($\gamma_{ge} \gg \gamma_{0e}$) — the time-scales of anti-bunching and blinking do not overlap[84]. The states ψ_g and ψ_e quickly reach an equilibrium with each other between each transition to ψ_0. Since we are not interested in anti-bunching in this section, we approximate ρ_g and ρ_e with their equilibrium values in the absence of the dark state ψ_0, and combine ψ_g and ψ_e into a single effective bright state ψ_1. We determine the transition rate γ_{01} from ψ_1 to ψ_0 using ρ_e^∞ from Eq. 6.3 to get

$$\gamma_{01} = \frac{\gamma_{0e}\gamma_{eg}}{\gamma_{eg} + \gamma_{ge}}, \tag{6.42}$$

and the effective transition rate from ψ_0 back to ψ_1 remains just $\gamma_{10} = \gamma_{e0}$. As a consistency

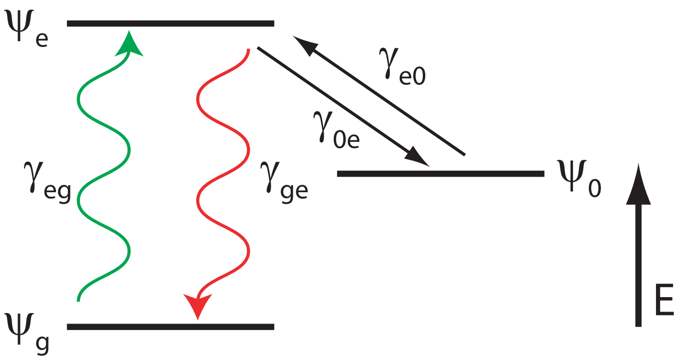


Figure 6.11: Simple three-state blinking model. A non-radiative transition from the excited state ψ_e to a dark state ψ_0 (occurring at rate γ_{0e}) renders the qdot off, and a non-radiative transition back to the excited state (at rate γ_{e0}) turns the qdot back on.

check, it is straightforward to show that this adiabatic elimination yields an identical stationary distribution to that of the full three-state model. It is important to note that the simplification to a two-state system is only valid at low excitation intensities, because an Auger process causes rapid transitions to the dark state when multiple electrons are excited simultaneously by high excitation intensities[71, 72]. This model can be improved by incorporation of a fourth, twice-excited state, but this prevents the simplification by adiabatic elimination. In our experiments, we will keep the excitation intensity relatively low so that we do not have to worry about these effects.

We now determine the effect of blinking on the FCS curve. We treat the blinking as a discrete modulation of the fluorescence quantum yield Φ between the values 0 and 1, independent of any other process determining the fluorescence rate. Following the procedure described in Section 4.1.3 we can incorporate blinking into the FCS curve by computing the correlation function of the blinking process on its own and multiplying the result by the FCS curve for the other dynamics in the experiment, such as systematic tracking errors. The blinking process that we have described here is known as a *random telegraph process*[8], and is different than the point processes we have dealt with when computing correlation functions elsewhere in this thesis. The f_n do not exist for this type of process, but we define the correlation function in a manner consistent with our previous definition:

$$g_2^b(\tau) = \frac{\langle \Phi^t \Phi^{t+\tau} \rangle}{\langle \Phi^t \rangle \langle \Phi^{t+\tau} \rangle} - 1, \tag{6.43}$$

where the superscript b indicates this is the correlation function of the blinking process. Since only the population of the bright state contributes to these averages, the steady-state value for the quantum yield is

$$\left\langle \Phi^t \right\rangle = \rho_1^\infty = \frac{\gamma_{10}}{\gamma_{01} + \gamma_{10}}. \tag{6.44}$$

The time correlation may be computed in a fairly simple manner. We write an equation for the evolution of ρ_1^t over the infinitesimal time step $\mathrm{d}t$:

$$\rho_1^{t+\mathrm{d}t} = \rho_1^t \left(1 - \gamma_{01}\mathrm{d}t\right) + \left(1 - \rho_1^t\right)\gamma_{10}\mathrm{d}t, \tag{6.45}$$

where the first term on the right is the probability that $\Phi^t = 1$ and that no transition to the dark state was made during the time step, and the second term is the probability that $\Phi^t = 0$ and that a transition to the bright state occurred during the time step. Rearranging terms in Eq. 6.45 and taking the limit $\mathrm{d}t \to 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\rho_1^t = -\left(\gamma_{01} + \gamma_{10}\right)\rho_1^t + \gamma_{10}. \tag{6.46}$$

We find the conditional probability $p(t + \tau | t)$ that $\Phi^{t+\tau} = 1$ *given* $\Phi^t = 1$ by solving Eq. 6.46 with the initial condition $\rho_1^t = 1$, and we use it to compute the correlation function

$$g_2^b(\tau) = \frac{\gamma_{01}}{\gamma_{10}} e^{-(\gamma_{01}+\gamma_{10})\tau}. \tag{6.47}$$

This expression decays exponentially on a characteristic time-scale $\tau_b = \left(\gamma_{01} + \gamma_{10}\right)^{-1}$. The contrast $g_2^b(0) = \gamma_{01}/\gamma_{10}$ is related to the fraction of time that the qdot spends in the bright state by $g_2^b(0) = \left(\rho_1^\infty\right)^{-1} - 1$, so that the more time the qdot spends in the dark state, the larger the blinking contribution to the FCS curve becomes.

The three-state system is a fair first approximation to qdot blinking behavior. However, the exponential decay it predicts in the correlation function does not generally match the statistics observed in real qdots, which have on- and off-times that are distributed with power law statistics[73, 74]. The addition of more dark states would introduce more exponential decays, and could therefore be used to approximate the observed power-law statistics[88]. However, it is beyond the scope of our work to try to do this and no physical insight would be gained by the indiscriminate addition of states. The physically-motivated explanation of the power-law statistics of qdot blinking is currently a topic of active research in theoretical chemistry groups[89].

6.4.2 Tracking-FCS of blinking quantum dots

We measured blinking on the qdots we purchased from Invitrogen by tracking them and studying their fluorescence correlation functions over longer time-scales than those discussed in

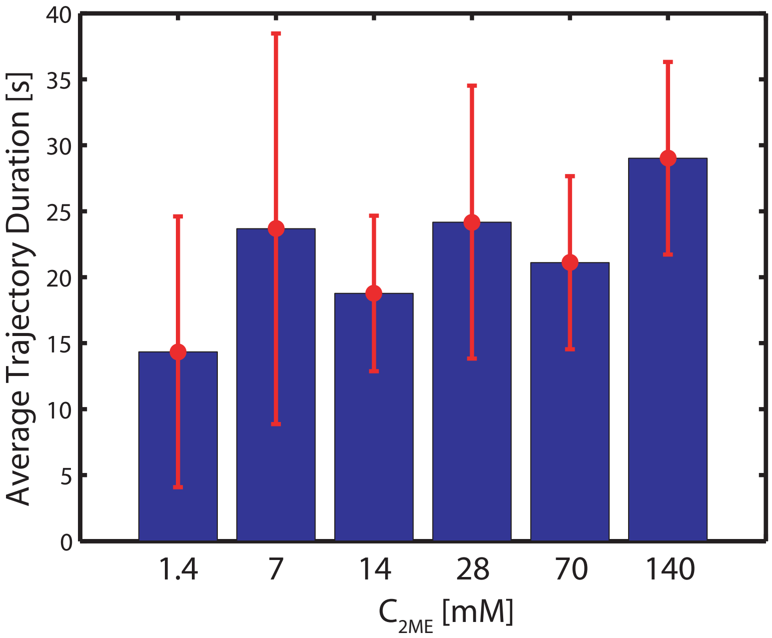


Figure 6.12: Average tracking trajectory durations at each $C_{2\mathrm{ME}}$, with 2σ error bars.

Section 6.3. We prepared 80fM solutions of qdots in an aqueous buffer consisting of 30mM sodium borate and 40% v/v glycerol. The buffer pH was adjusted to 8.3, after the addition of glycerol, by titration with sodium hydroxide. 2-mercaptoethanol was added at six different concentrations $C_{2\mathrm{ME}}$ immediately before the addition of qdots to the buffer. Qdot solutions were diluted to 80fM to prevent multiple qdots from being detected over the course of a single tracking trajectory. At all $C_{2\mathrm{ME}}$, we fixed the excitation intensity to 570W/cm^2 using an absorptive neutral density filter.

We tracked between 10 and 13 qdots at each 2-mercaptoethanol concentration. The average tracking trajectory durations at each $C_{2\mathrm{ME}}$ are shown in Fig. 6.12. There may be a slight trend toward increased trajectory duration with increased $C_{2\mathrm{ME}}$, but this trend is not larger than the experimental uncertainty and so is not statistically significant. This indicates that $C_{2\mathrm{ME}}$ was never low enough for a largely increased blinking time to end many trajectories prematurely. These results agree with previous observations[76] that found nearly complete suppression of qdot blinking for $C_{2\mathrm{ME}} > 1.4$mM on time-scales longer than 1s.

We computed $g_2(\tau)$ for each tracked qdot, and averaged all of the curves from each 2-mercaptoethanol concentration. The results are shown in Fig. 6.13. In the figure we also provide the $g_2(\tau)$ curve for a 60nm fluorescent bead (Bangs Laboratories) in water, which had nearly the same diffusion coefficient as the qdots in 40% v/v glycerol and was excited at a laser intensity that produced a nearly identical fluorescence rate to that of the qdots. By taking the mean-squared deviation of the tracking stage positions, we determined the RMS tracking

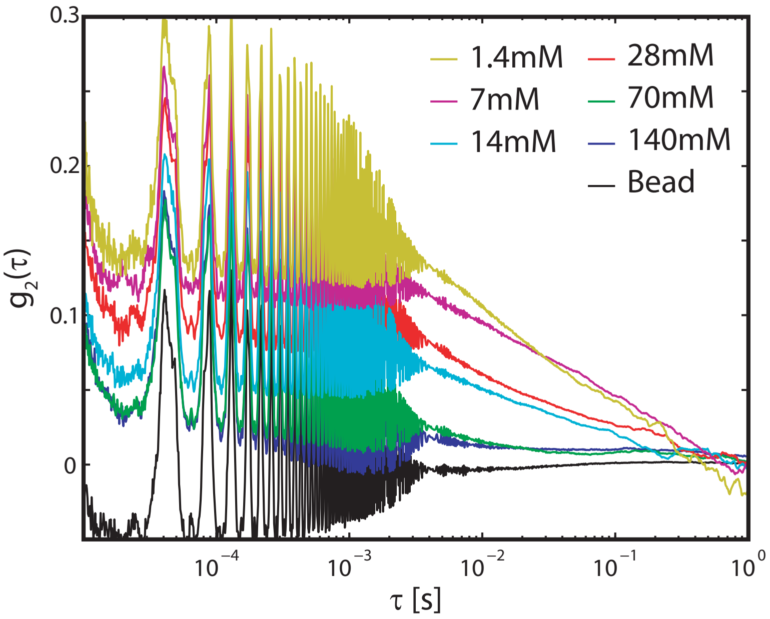


Figure 6.13: Tracking-FCS curves for qdots in 40% v/v glycerol/water solution with varied $C_{2\mathrm{ME}}$, compared to a 60nm fluorescent bead in water.

error of the bead and qdots to be between 150nm and 200nm along the x and y axes and between 200nm and 350nm along the z axis. As a consequence of these small errors relative to the 1.2μm and 1.4μm beam waists, we see an offset $g_2(0)$ of essentially zero in the FCS curve for the bead. For the qdot curves, we attribute the increasing $g_2(0)$ as $C_{2\mathrm{ME}}$ decreases to reduced suppresion of qdot blinking. The FCS curves for the qdots at all $C_{2\mathrm{ME}}$ have decayed to zero by $\tau \approx 1$s, indicating that we observe blinking only on times shorter than this. This is partly because longer off-times are not tolerated by the apparatus: the qdot may move too far while off for it to be detected once it turns on. These $g_2(\tau)$ therefore represent qdot FCS curves *conditioned* on never switching off for too long. We expect this contribution to our data to be small because we were able to track qdots for long periods at all 2-mercaptoethanol concentrations; however, this effect can be eliminated in the future by labeling the qdots with organic dyes as done in [75], allowing us to probe blinking on even longer time-scales.

We first use the three-state model described in the previous section to try to fit the decays in the FCS curves. We used the function

$$g_2(\tau) = \left[1 + a + \frac{\gamma_{01}}{\gamma_{10}} e^{-(\gamma_{01}+\gamma_{10})\tau}\right] \left[g_2^0(\tau) + 1\right], \tag{6.48}$$

derived from Eq. 6.47, where we inserted the experimentally-measured curve for the fluorescent bead into $g_2^0(\tau)$, and we included the parameter a to accommodate a slight offset in the $g_2(\tau)$ curves over time-scales shorter than 1s that we suspect is a systematic artifact due to slow

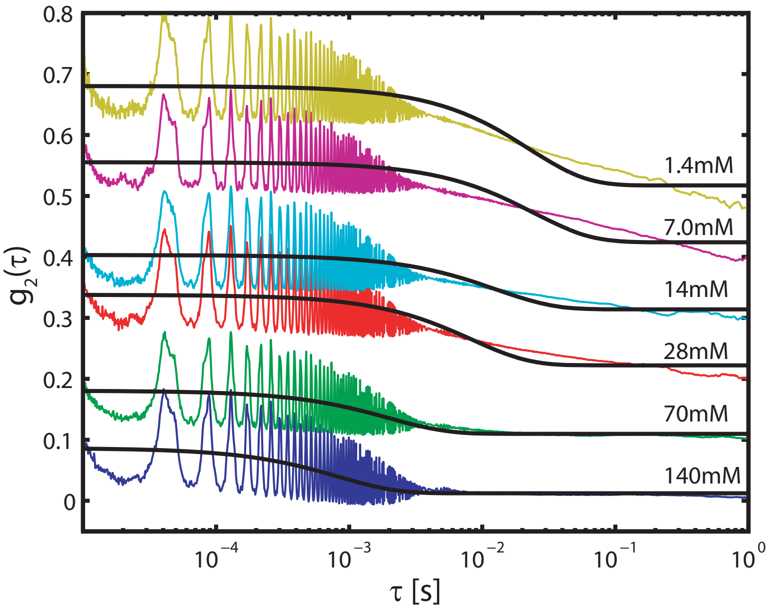


Figure 6.14: Blinking FCS curves from Fig. 6.13 with fits to the exponential curve in Eq. 6.48. The curves are spaced apart by 0.1 for clarity. Fit parameters are given in Table 6.1

C_{2ME} [mM]	γ_{10} [s^{-1}]	γ_{01} [s^{-1}]	$a \cdot 10^3$	τ_b[ms]	ρ_1^∞
1.4	38	6.2	17	23	0.86
7.0	36	4.8	24	24	0.88
14	72	6	14	13	0.92
28	110	12	22	8.3	0.90
70	470	33	1.0	2.0	0.93
140	1100	83	12	0.83	0.93

Table 6.1: Parameters for the fits in Fig. 6.14

fluctuations in the qdot's fluorescence rate. The fits of Eq. 6.48 to the data are shown in Fig. 6.14 and the fit parameters are shown in Table 6.1. The exponential curves in Fig. 6.14 fit the rather steep correlation decays in the data at the higher 2-mercaptoethanol concentrations ($C_{2ME} \geq 70$mM) quite well. This implies that the blinking statistics at those concentrations are roughly exponential, which is consistent with the hypothesis that high concentrations of 2-mercaptoethanol reduce the number of dark states available for the excited electrons within the qdot[76]. At lower C_{2ME}, however, the correlation functions decay much less sharply and the result is increasingly poorer fits as C_{2ME} decreases. This deviation from a single exponential decay cannot result from a single dark state with transition statistics governed by only a single set of forward and backward rates.

In order to better fit the correlation decay curves at lower C_{2ME}, we require a more detailed model of the qdot blinking dynamics. As we discussed in the previous section, however, the physics of qdot blinking are still not fully understood so no model provides an obviously appropriate choice. Experimental observations have shown that a power law provides a good

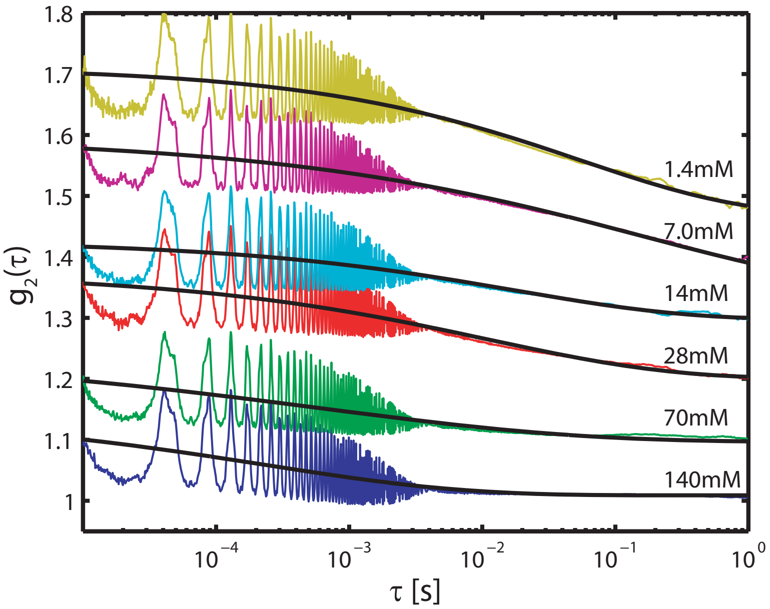


Figure 6.15: Blinking FCS curves from Fig. 6.13 with fits to the stretched exponential curve in Eq. 6.49. The curves are spaced apart by 0.1 for clarity. Fit parameters are given in Table 6.2

empirical description of qdot blinking[73, 74], but it is difficult to compute the correlation function for power law statistics. A correlation function may be found for qdots with power law distributed off-times if we introduce time-scale cutoffs and require the qdot on-times to be exponentially distributed[90], but these cutoffs are not physical and the exponential on-times do not agree with observation[73, 74, 89]. Instead, we let our data motivate our choice of fitting function and leave a more detailed understanding to future work or collaboration.

We find that the introduction of the the stretching exponent β to Eq. 6.48, giving the function

$$g_2(\tau) = \left[1 + a + \frac{\gamma_{01}}{\gamma_{10}} e^{-(\gamma_{01}+\gamma_{10})^{\beta}\tau}\right]\left[g_2^0(\tau) + 1\right], \tag{6.49}$$

results in curves that fit the measured correlation functions at low $C_{2\mathrm{ME}}$ very well. These fits are shown in Fig. 6.15 and the fit parameters are given in Table 6.2. At the lower concentrations $C_{2\mathrm{ME}} \geq 70mM$, the steep decays in the data may cause artificially high γ values for these fits. As a result, trends in those parameters may be exaggerated over these concentrations.

Despite its empirical nature, we chose to leave Eq. 6.49 in a form suggestive of Eq. 6.48 because the parameters γ_{10} and γ_{01} share some physical relevance in the two equations. In both, $\tau_b = (\gamma_{01} + \gamma_{10})^{-1}$ represents a characteristic time-scale of correlation decay and γ_{01}/γ_{10} is proportional to the the variance in the fluorescence quantum yield due to blinking and can be used to find the equilibrium population of the bright state ρ_1^{∞} as described in the previous section. We use the fitted values of γ_{01} and γ_{10} to find τ_b and ρ_1^{∞}, and we provide these values in

C_{2ME} [mM]	γ_{10} [s^{-1}]	γ_{01} [s^{-1}]	β	$a \cdot 10^3$	τ_b[ms]	ρ_1^∞
1.4	15	3.6	0.36	30	55	0.81
7.0	4.9	1.3	0.26	61	161	0.79
14	50	6.4	0.34	2.5	18	0.89
28	76	13	0.32	0.58	11	0.85
70	1400	210	0.22	3.3	0.62	0.87
140	3800	530	0.29	9.0	0.22	0.88

Table 6.2: Parameters for fits in Fig. 6.15.

Tables 6.1 and 6.2. While we may directly compare values for the γ parameters between fits to the exponential model, the stretched exponential form generally prohibits such comparisons for any curves with different β values. In our fits, however, the values for β lie fairly close together — the average over all concentrations is $\beta = 0.30 \pm 0.04$ — so coarse comparisons may be made.

Both fits suggest the same general trend, in agreement with measurements on surface-immobilized qdots on time-scales longer than 1s[76]: with increasing C_{2ME}, the fraction of time the qdot spends in the bright state ρ_1^∞ increases. Those prior measurements suggested that the total on-time increased because the average individual on-time increased while the average individual off-time remained constant. Both of our sets of fit parameters suggest that this was not the case for the short time-scale blinking that we measured. The derivative

$$\frac{\mathrm{d}}{\mathrm{d}\gamma_{01}}\tau_b = -\frac{1}{(\gamma_{01} + \gamma_{10})^2} \tag{6.50}$$

is always negative so that any increases in on-time (decreases in γ_{01}), while off-times remain constant, will be met by increases in τ_b. Contrary to previous measurements, we observe that the correlation times τ_b actually *decrease* as C_{2ME} increases. This implies that any increases in on-times due to increased C_{2ME} must be offset by decreases in off-times. Without considering the individual fits to the γ parameters, we cannot rule out the possibility that both on-times and off-times *decrease* with increasing C_{2ME}, but do so in proportion to each other so that the average total on-time increases. In fact, all of our fitted γ values suggest that this is the case — we must hesitate to make this claim with certainty, however, because of the difficulty in interpreting the different β values resulting from the fits to the stretched exponential.

Chapter 7

Structural fluctuations in isolated DNA molecules

The final application of the tracking apparatus that we discuss in this thesis is to the study of internal structural fluctuations of double-stranded λ-phage DNA molecules. While this is the first application of our work to a real biological system, our measurements are not yet directly relevant to the biological properties of these molecules. Instead we study them strictly from a polymer dynamics perspective as we attempt to reconcile our measurements with measurements made with other techniques and with predictions from polymer theory.

Figure 7.1 provides a basic illustration of the molecules that we study in this chapter. DNA molecules are labeled at random sites along their backbones by intercalating organic dyes. Some of these dyes absorb green light and emit red fluorescence; these are used as tracers for the tracking system to follow the molecule. The remaining dyes absorb violet light and emit blue fluorescence, and they are excited by the 415nm probe laser. The probe's small focused beam waist leads to large fluctuations in the blue fluorescence signal as the dyes drift through the beam. Motion of the dyes relative to the probe beam is caused almost exclusively by structural fluctuations within the DNA molecule because the tracking system eliminates contributions due to the molecule's translational motion. By studying the statistics of the blue fluorescence, we may infer basic mechanical properties of the freely-diffusing DNA molecule.

Polymers are very complex molecules. The theory used to describe them is necessarily approximate, because the exact equations of motion for the polymer would be impossible to solve in any analytical way. While these approximate theories rely on free parameters that are best determined by experiment, they have yielded some predictions — for example, the scaling of the polymer's center of mass diffusion coefficient or radius of gyration with respect to the chain length — that agree with experimental observations to a high degree of accuracy. In this chapter we study the predictions of applicable theories to the dynamic behavior of DNA molecules; we study both center of mass motion and internal motion, which are almost

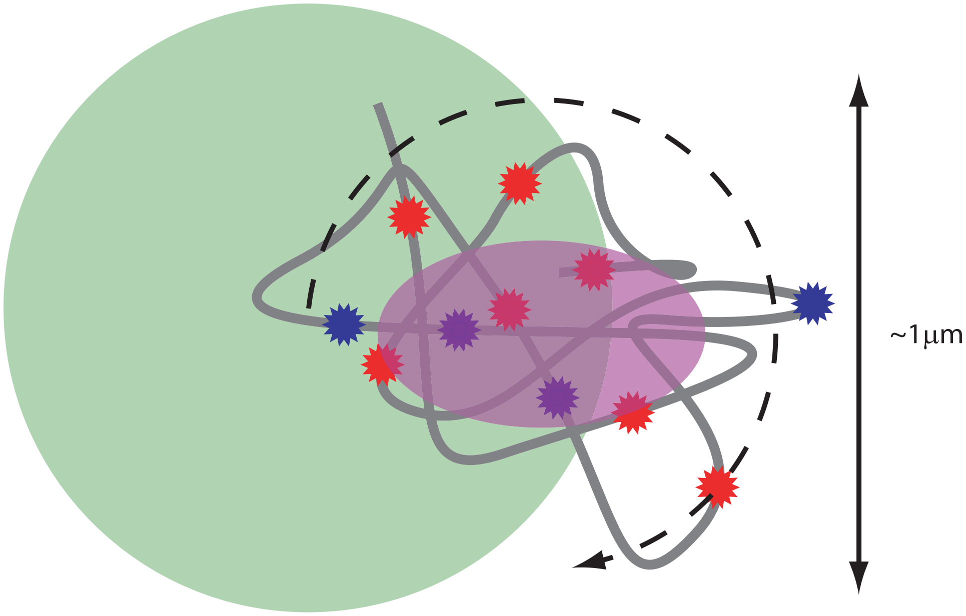


Figure 7.1: Illustration of a DNA molecule (gray) labeled with intercalating dyes. The rotating tracking beam (green) and probe beam (violet) are drawn approximately to scale.

completely decoupled in our measurements.

Recent literature from several groups[21–25] has reported on measurements of a similar nature to ours. However, despite these measurements a fair amount of uncertainty still remains about some basic polymer properties of DNA molecules. Surprisingly, there is no consensus in the literature on the value of the free-solution diffusion coefficient of λ-phage DNA, with measured values differing by as much as a factor of 3; some of these differences have been accounted for, some have not. More interestingly, there is some uncertainty regarding the basic physics governing the internal polymer motion: the importance of the interactions between the polymer and its solvent is still not agreed upon. Matters are certainly not simplified by theoretical deficiencies that we find in some of those references. We address all of the concerns regarding this literature in Section 7.3 of this chapter.

This chapter begins with a detailed overview of the aspects of polymer theory relevant to our work. This review is based closely on the definitive book by Doi and Edwards[91], but for completeness we provide enough detail in the review to accommodate all of the calculations that we will need for computing FCS curves later in the chapter. Once those calculations are complete, we review the recent literature relevant to our work in the context of the polymer theory. Finally we present our measurements, in which we directly measure the translational diffusion of the molecule and find a center of mass diffusion coefficient that lies about half-way between previous measurements. In addition, we clearly detect statistical correlations in the molecule's fluorescence on time-scales up to 1 second, but we find the measurements to lack any clear signatures of either the presence or absence of strong solvent interactions.

7.1 Theoretical dynamics of linear polymers

A *polymer* is any molecule constructed by linking together a large number of similar or identical smaller molecules or *monomers*. Three of the four major classes of biological molecules — proteins, nucleic acids and carbohydrates, but not lipids — are examples of polymers. Due to their large size, a complete specification of even the static properties of a single polymer would require accounting for an extremely large number of degrees of freedom: the λ-phage DNA molecules that we study in this chapter are composed of $\sim 10^6$ atoms that are each free to move in three dimensions and to interact with surrounding solvent molecules.

A half-century of experimental evidence suggests that many properties of polymers are characteristics not of the specific form of individual monomers, but rather of the fact that the monomers are linked together to form long chains[92]. These so-called *global* properties can be described in terms of only a few adjustable parameters that are specific to individual polymers. It is the goal of polymer physics to find these general theoretical descriptions, and the goal of polymer chemistry to understand the localized behaviors that determine the global parameters. In this section we review the basic theory describing the mechanical properties of flexible linear polymers from the global polymer physics perspective. Most of this review is drawn from the definitive book by Doi and Edwards[91], with some ideas taken from de Gennes' book[92]. Both are excellent resources, the latter as a more readable and insightful introduction and the former for its detailed calculations.

7.1.1 Static properties of flexible polymers

We begin by introducing a basic model that will form the foundation of many of our calculations in this chapter. We let the polymer molecule consist of a sequence of N submolecules that are assembled into a linear chain. These submolecules may correspond to the individual monomers, but our definition is sufficiently general that it may likewise apply to a coarser description where each submolecule consists of a group of monomers (and b is chosen accordingly). The fact that this choice is not well-defined is not a problem with our model, because coarse-graining the polymer into submolecules is itself an artificial description and will yield incorrect local properties for absolutely any choice of submolecules. On a large enough scale, however, we will find that the specific assignment of submolecule size does not affect the predictions for the polymer's global behavior.

We let the vector $\mathbf{R}_m$ denote the position of the m^{th} submolecule relative to an arbitrary

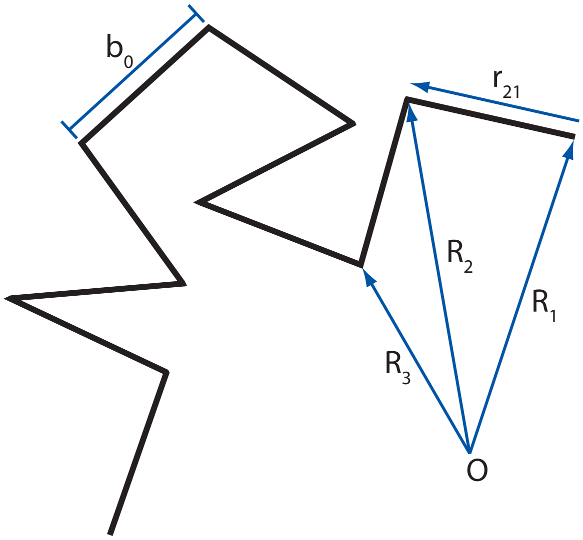


Figure 7.2: Freely-jointed polymer model. The vectors $\mathbf{R}_1$, $\mathbf{R}_2$ and $\mathbf{R}_3$ to the first three submolecules (from arbitrary origin O) are shown, as is the vector $\mathbf{r}_{21} \equiv \mathbf{R}_2 - \mathbf{R}_1$.

coordinate origin. A simple model of the polymer requires that the length of the bonds between submolecules be fixed at b_0 and allows the submolecules to rotate freely about their bonds in three dimensions. This is known as the *freely-jointed* chain and is illustrated in Fig. 7.2. In solution, each submolecule experiences a random Brownian force (see Section 3.1.1) from collisions with the solvent molecules. As a result we treat the instantaneous configuration of the polymer as a random walk starting from $\mathbf{R}_1$ and proceeding according to the prescribed behavior of the bonds between submolecules. We define the vector from submolecule m to submolecule n by $\mathbf{r}_{nm} = \mathbf{R}_n - \mathbf{R}_m$. We compute the statistics of $\mathbf{r}_{nm}$, for $n \geq m$, by

$$\langle \mathbf{r}_{nm} \rangle = \left\langle \sum_{j=1}^{n-m} \mathbf{r}_{(m+j)(m+j-1)} \right\rangle = 0 \tag{7.1}$$

$$\left\langle |\mathbf{r}_{nm}|^2 \right\rangle = \left\langle \sum_{j=1}^{n-m} \sum_{k=1}^{n-m} \mathbf{r}^T_{(m+j)(m+j-1)} \mathbf{r}_{(m+k)(m+k-1)} \right\rangle = (n-m) b_0^2, \tag{7.2}$$

where the second result follows from the fact that $\left\langle \mathbf{r}^T_{nm} \mathbf{r}_{jk} \right\rangle = \delta_{nj}\delta_{mk} b_0^2$ because the vectors between submolecules are uncorrelated. By symmetry, we may generalize this result to arbitrary n, m: $\left\langle |\mathbf{r}_{nm}|^2 \right\rangle = |n-m| b_0^2$. In fact, for large $|n-m|$ it can be shown that the *distribution* of $\mathbf{r}_{nm}$ is approximately Gaussian with mean and variance given by Eqs. 7.1 and 7.2[91].

Polymer models with rigid bond lengths, of which the freely-jointed chain is an example, are known as *random-flight* models. A broad range of these types of models can be imagined; for example, a *freely-rotating* chain in which the angle between any three sequential submolecules is fixed but each submolecule is permitted to rotate about its bonds is both physically sensible and solvable for the first two moments of $\mathbf{r}_{nm}$. Interestingly, these models — which differ in their local structure — result in nearly identical statistics to the freely-jointed chain for $\mathbf{r}_{nm}$ on global scales (when $|n - m|$ is large)[91]. The only difference is generally just the replacement of b_0 by an effective bond length $b > b_0$.

That the local properties of the random-flight models do not affect the Gaussian statistics of $\mathbf{r}_{nm}$ over long distances suggests a further simplification of these models. We already discussed the arbitrary nature of our choice of submolecules; we may as well choose our submolecules to be large enough that the local stiffness of the bonds between them may be neglected from our model. For example, suppose a zero-mean Gaussian with variance $N_0 b_0^2$ is a good approximation to the statistics of $\mathbf{r}_{nm}$ for $|n - m| \geq N_0$ in a random-flight model. We further coarsen this model into a sequence of larger submolecules, each containing at least N_0 of the random-flight submolecules, and let the effective bond length $b = b_0\sqrt{N_0}$. The resulting statistics of the vectors between the n^{th} and m^{th} of these new submolecules are zero-mean Gaussian with variance $|n - m|b^2$. This simplified model is known as the Gaussian chain, and is very important because of its analytical simplicity.

We must always remember what physical properties of the polymer we are choosing to ignore through the selection of our polymer model. In the case of the Gaussian chain, local chain stiffness is eliminated; any polymer that is stiff over length-scales comparable to either the length of the polymer chain or the beam waist of the excitation laser will exhibit signatures of this stiffness in its FCS curves[23]. For our purposes, however, the Gaussian chain model should suffice, because our relatively long DNA molecules do not satisfy these criteria and may be described as effectively flexible chains[23].

Another inaccuracy is that the Gaussian model describes polymer chains that are not prevented from intersecting themselves. Such "phantom chains" do not exist in the real world, where knots and tangles form within polymers much like in macroscopic strings as a consequence of self-intersection. Under special solution conditions known as Θ conditions, the repulsive force from self-interaction is exactly balanced by the interaction with the solvent, so that the equilibrium configuration of the real chain is no different than of the phantom chain[91, 93]. In more general conditions, known as "good solvents," the polymer swells due to self-avoidance and this phenomenon is difficult to model exactly in a dynamic context without producing unsolvable equations of motion.

7.1.2 Generalized dynamics of polymer solutions

Using the very basic description of a polymer as a sequence of linked submolecules, we may incorporate the motion of the chain into our theoretical description in a very general way. A polymer chain in solution moves because its backbone makes frequent collisions with surrounding solvent molecules that exert impulsive forces on it. The resulting polymer motion is just a generalization of the Brownian motion of a solid particle that we reviewed in Section 3.1.1. In fact, we describe the Brownian force acting on the polymer in terms of the forces acting on each submolecule, and we set the statistics of these forces to exactly what they would be if the submolecules were freely-diffusing particles independent of the polymer chain.

Adjacent submolecules in the polymer chain model exert forces on each other that hold the polymer together. We may capture all of these forces in terms of a generic potential U that depends on the polymer's instantaneous configuration. One important aspect of polymer Brownian motion that is distinct from that of a single solid particle is that the motion of individual submolecules is coupled through interactions with the solvent. Motion of one submolecule induces fluid flow that biases the motion of other submolecules that are nearby in space but potentially quite distant along the polymer chain. We use a self-adjoint positive definite interaction tensor $\mathbf{H}_{nm}$ (or *mobility* tensor) that maps the force applied on submolecule m to the velocity of submolecule n based upon this hydrodynamic coupling.

In order to express the motion of the chain with a differential equation, we calculate the force acting on every submolecule m in the polymer chain due to the confining potential U and the Brownian force $\mathbf{B}_m(t)$. These forces couple into the velocity of submolecule n according to $\mathbf{H}_{nm}$. The Langevin equation of motion for the submolecules of the polymer is then given by[91]:

$$\frac{\partial}{\partial t}\mathbf{R}_n = \sum_m \mathbf{H}_{nm} \cdot [-\nabla_m U + \mathbf{B}_m(t)] + \frac{k_B T}{2} \sum_m \nabla_m \cdot \mathbf{H}_{nm}, \tag{7.3}$$

where ∇_m is the del operation with respect to the components of the vector $\mathbf{R}_m$. The first term in Eq. 7.3 is due to the forces we discussed, and the second is a correction necessary whenever $\mathbf{H}_{nm}$ depends explicitly on the coordinates $\mathbf{R}_n$[94]. $\mathbf{B}_m(t)$ is simply a three-dimensional vector of random forces of the type described in Section 3.1.1: each component of $\mathbf{B}_m(t)$ is a zero-mean Gaussian white noise process. Due to the coupling of submolecules through the hydrodynamic interaction, the correlation function of the Brownian force is

$$\left\langle \mathbf{B}_n(t)\mathbf{B}_m(t')^T \right\rangle = 2k_B T\delta(t-t') \left(\mathbf{H}^{-1}\right)_{nm} \mathrm{Id}_3, \tag{7.4}$$

where $\mathbf{H}^{-1}$ is the inverse of the hydrodynamic tensor matrix.

It is generally fairly easy to compute reasonable U and $\mathbf{H}_{nm}$. Essentially, this reduces all of polymer dynamics to the solution of the Langevin equation Eq. 7.3 or its associated Fokker-Planck equation. Unfortunately, an explicit solution is only possible for a very small class of U and $\mathbf{H}_{nm}$. The trick now is to find a U and $\mathbf{H}_{nm}$ that are simple enough to yield exact solutions but detailed enough to account for as much of the relevant polymer physics as possible.

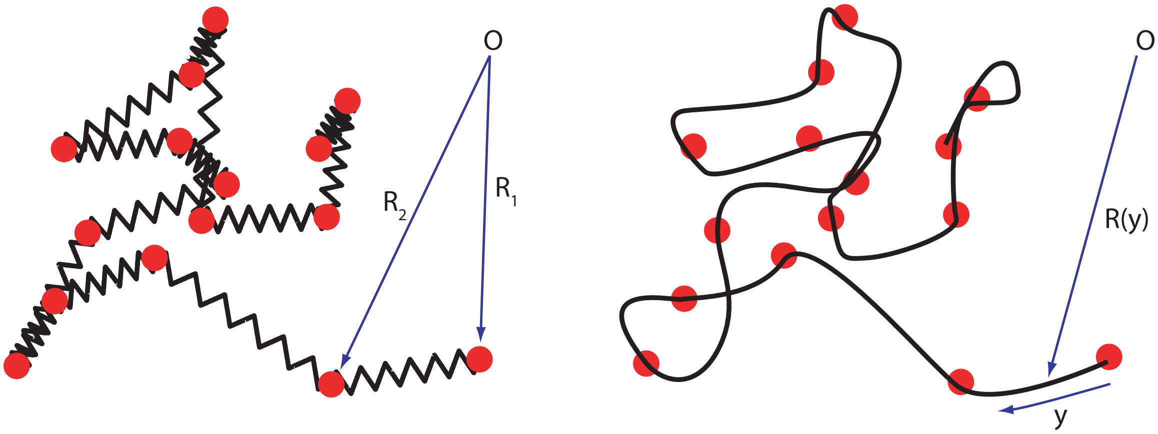


Figure 7.3: Rouse/Zimm model. **Left**: Adjacent submolecules are constrained by a Harmonic potential equivalent to being bound together by springs. **Right**: Continuum limit.

7.1.3 Harmonically-bound submolecules: the Rouse model

The first model to incorporate dynamics into the model of the Gaussian polymer chain was proposed by Rouse in 1953[95]. This model simplifies the U and $\mathbf{H}_{nm}$ to perhaps the greatest extent possible while maintaining the correspondence of the Langevin equation's steady-state solution with the static Gaussian chain. This model is considered insufficient for describing experimentally-observable properties of real polymers because it only incorporates the influence of localized interactions along the polymer chain. However, it is important because it forms the basis of more detailed models, and the equations of motion for that model reduce to a form resembling those for this model.

We let the N submolecules of the polymer be bound by the harmonic potential

$$U = \frac{k}{2} \sum_{n=2}^{N} |\mathbf{R}_n - \mathbf{R}_{n-1}|^2 , \tag{7.5}$$

where k is the force constant of the bonds between submolecules. We will deduce the value of k at the end of this section by equating the variance of the inter-chain distances predicted by this model with those of the static chains in Section 7.1.1. This potential describes the polymer as essentially a set of beads connected by springs, as illustrated in Fig. 7.3. The Rouse model ignores all hydrodynamic couplings between submolecules, so that the mobility tensor is given by

$$\mathbf{H}_{nm} = \frac{1}{\gamma} \mathrm{Id}_3 \, \delta_{nm}, \tag{7.6}$$

where γ is the Stokes drag coefficient of the individual submolecules, identical to that described in Section 3.1.1.

Inserting Eqs. 7.5 and 7.6 into Eq. 7.3, we get the Rouse Langevin equation

$$\gamma \frac{\mathrm{d}}{\mathrm{d}t}\mathbf{R}_n = -k\left(2\mathbf{R}_n - \mathbf{R}_{n+1} - \mathbf{R}_{n-1}\right) + \mathbf{B}_n(t). \tag{7.7}$$

At the ends of the chain, we just use $R_0 = R_{N+1} = 0$. Instead of solving this (potentially high-dimensional) ordinary differential equation, we make the assumption that the number N of submolecules is very large, so that b is very small compared to the overall length of the polymer. In this case we may approximate the polymer as a continuous curve parametrized by the vector $\mathbf{R}(y,t)$, where $y = nb$ is the position coordinate along the backbone of the chain. This continuum limit is illustrated in Fig. 7.3, and results in the transformation of Eq. 7.7 into the partial differential equation

$$\gamma \frac{\partial}{\partial t}\mathbf{R}(y,t) = kb^2 \frac{\partial^2}{\partial y^2}\mathbf{R}(y,t) + \mathbf{B}(y,t). \tag{7.8}$$

The statistics of $\mathbf{B}$ are transformed using $\delta_{nm} \to b\delta(y_n - y_m)$, so that

$$\begin{aligned} \langle \mathbf{B}(y,t)\rangle &= 0 \\ \left\langle \mathbf{B}(y,t)\mathbf{B}(y',t')^T \right\rangle &= 2k_B T \gamma b \delta(y - y')\delta(t - t')\,\mathrm{Id}_3, \end{aligned} \tag{7.9}$$

and the boundary conditions for $\mathbf{R}$ simply become

$$\left.\frac{\partial}{\partial y}\mathbf{R}(y,t)\right|_{y=0,y=L} = 0. \tag{7.10}$$

Our definition of the continuum limit is slightly different than that used in[91], where the variable n is simply turned into a continuous variable ranging between 0 and N. Such an n is dimensionless while y is not, which results in the factor of b in the covariance in Eq. 7.9. The choice of how to define the limit is not important because it ultimately must not affect the statistics we compute, but we prefer our definition because it is at least mathematically well-defined, corresponding to taking $N \to \infty$, $b \to 0$ and $Nb = L$. Based on this definition we know that k must contain a b^{-2} term in order to ensure that the second derivative term is nonzero, but we don't worry about this until we can compute k explicitly.

Due to these boundary conditions and the presence of the second derivative operator, we project Eq. 7.8 into the Fourier basis by defining

$$\chi_p(t) = \frac{1}{L}\int_0^L \mathrm{d}y \mathbf{R}(y,t)\cos\left[\frac{\pi p y}{L}\right] \tag{7.11}$$

$$\Lambda_p(t) = \left(2 - \delta_{0p}\right)\int_0^L \mathrm{d}y \mathbf{B}(y,t)\cos\left[\frac{\pi p y}{L}\right] \tag{7.12}$$

for $p \in \mathbb{N}$. It is important to note that these definitions are not normalized (χ_p is missing a factor of $2 - \delta_{0p}$, while Λ_p is missing $1/L$), which matters when we transform back into real-space coordinates. The Langevin equation for the Fourier coefficients is

$$\gamma_p \frac{\partial}{\partial t} \chi_p(t) = -k_p \chi_p(t) + \Lambda_p(t), \tag{7.13}$$

where we have defined $k_p = 2\pi^2 k p^2 b^2 / L$ and $\gamma_p = (2 - \delta_{0p})\gamma L$. The statistics of $\Lambda_p(t)$ are found directly from Eq. 7.9:

$$\left\langle \Lambda_p(t) \Lambda_q(t')^T \right\rangle = 2 k_B T \gamma_p b \delta_{pq} \delta(t - t') \, \mathrm{Id}_3 \,. \tag{7.14}$$

We now solve Eq. 7.13 to get

$$\chi_p(t) = \exp\left[-\frac{k_p}{\gamma_p} t\right] \chi_p(0) + \frac{1}{\gamma_p} \int_0^t \mathrm{d}\tau \exp\left[-\frac{k_p}{\gamma_p}(t - \tau)\right] \Lambda_p(\tau), \tag{7.15}$$

and use this expression to compute the statistics of χ_p. For $p > 0$ $\chi_p(t)$ is stationary, so we compute its statistics by taking the limit $t \to \infty$. This gives

$$\begin{aligned} \left\langle \chi_p(t) \right\rangle &= 0 \\ \left\langle\left\langle \chi_p(t + \tau) \chi_q(t)^T \right\rangle\right\rangle &= \frac{k_B T b}{k_p} \exp\left[-\frac{k_p}{\gamma_p} \tau\right] \delta_{pq} \, \mathrm{Id}_3, \end{aligned} \tag{7.16}$$

(where we use van Kampen's covariance notation[8], $\langle\langle \cdot \rangle\rangle$, as in Chapter 3) indicating that the Fourier coefficients decay exponentially on time-scales that depend inversely as the square of the wave number p. Low-frequency modes therefore dominate the predicted behavior of the polymer. For $p = 0$, the exponential terms in Eq. 7.15 vanish because $k_0 = 0$ and we are left with

$$\begin{aligned} \langle \chi_0(t) \rangle &= \chi_0(0) \\ \left\langle\left\langle \chi_0(t + \tau) \chi_0(t)^T \right\rangle\right\rangle &= \frac{2 k_B T b}{\gamma_0} t \, \mathrm{Id}_3, \end{aligned} \tag{7.17}$$

which clearly indicates that χ_0 is nonstationary.

We require the statistics of the polymer chain's motion in position space, not Fourier space. We invert the projection into the Fourier basis by computing

$$\mathbf{R}(y, t) = \chi_0(t) + 2 \sum_{p=1}^{\infty} \chi_p(t) \cos\left[\frac{\pi p y}{L}\right], \tag{7.18}$$

and we use this and Eqs. 7.16 and 7.17 to find the statistics of the polymer's motion. In particular, the center of mass coordinate is given by

$$R_{cm}(t) = \frac{1}{L}\int_0^L \mathrm{d}y\mathbf{R}(y,t) = \chi_0(t), \tag{7.19}$$

so comparison of Eqs. 3.8 and 7.17 indicates that the center of mass moves by ordinary Brownian motion with diffusion coefficient $D_g = k_BTb/\gamma L$. Furthermore, we can compute the statistics of the end-to-end vector

$$\mathbf{R}(L,t) - \mathbf{R}(0,t) = -4\sum_{p=1}^{\infty}\chi_{2p-1}(t), \tag{7.20}$$

which is stationary with zero mean and variance given by

$$\left\langle|\mathbf{R}(L,t) - \mathbf{R}(0,t)|^2\right\rangle = 16\sum_{p=1}^{\infty}\frac{3k_BTL}{2b\pi^2k(2p-1)^2} = \frac{3k_BTL}{bk}. \tag{7.21}$$

From our description of the static chain in Section 7.1.1, specifically Eq. 7.2, we must require $\left\langle|\mathbf{R}(L,t) - \mathbf{R}(0,t)|^2\right\rangle = Nb^2$. This tells us that $k = 3k_BT/b^2$ is the appropriate spring constant for the potential U binding the submolecules together, and allows us to compute the decay time-scale of mode p from Eq. 7.16,

$$\tau_p \equiv \frac{\gamma_p}{k_p} = \frac{\gamma L^2}{3k_BT\pi^2p^2} = \frac{\tau_1}{p^2}. \tag{7.22}$$

The predictions in Eqs. 7.19 and 7.22 are considered major failures of the Rouse model[91]. Empirical data has shown that D_g scales more closely to $L^{-1/2}$ and that τ_p scales as $L^{3/2}$; the deviations of the Rouse predictions from these values is attributed to the exclusion of hydrodynamic interactions from the model. We now review one way in which those interactions are incorporated while maintaining the linearity, and hence solubility, of the Langevin equation.

7.1.4 Hydrodynamic interactions: the Zimm model

In the previous section we offered a detailed derivation of the statistics of the Rouse polymer, despite the fact that the Rouse model is considered insufficient for describing real polymers. The reason for the deficiency is that the model ignores all interactions between the polymer and the solvent and describes a phantom chain, ignoring self-crossing of the polymer. The difficulty in incorporating these phenomena into the model is not a modeling difficulty, but rather a mathematical one: the general Langevin equation for the motion of the polymer becomes nonlinear when the appropriate mobility tensor $\mathbf{H}_{nm}$ and self-avoidance potential U_0 are used. However, a simplification of these exact terms attributed to Zimm[96] has improved some of the deficiencies of the Rouse model while remaining exactly solvable.

Zimm's approach used the Oseen tensor, which is derived straightforwardly from the stress tensor for the fluid[91, 97], as the mobility tensor. The Oseen tensor is nonlinear, but its divergence vanishes so that the correction term in Eq. 7.3 is zero. Zimm simplified the nonlinearity by what is known as a mean-field approximation: he replaced the Oseen tensor by its *average* over the ensemble of static polymer configurations, as described in Section 7.1.1. This results in a linear Langevin equation for $R(y,t)$.

Without providing the details because they are worked out in [91, 96], Zimm was able to reduce the Langevin equation for his polymer model to one of exactly the same form as that for the Rouse polymer,

$$\gamma_p^Z \frac{\partial}{\partial t}\chi_p(t) = -k_p^Z \chi_p(t) + \Lambda_p(t), \tag{7.23}$$

which differs from Eq. 7.13 only in the use of the Zimm values for the coefficients γ_p^Z and k_p^Z, which are given in [91]. This model, despite the use of the mean field and other approximations, predicts the diffusion coefficient and relaxation time scaling $D_g \sim L^{-1/2}$ and $\tau_1 \sim L^{3/2}$ discussed in the previous section.

So far the polymer we have described is still a phantom chain, so any predictions about it are only truly valid in Θ solution conditions. Self-avoidance can be incorporated into the polymer model by addition of a repulsive potential to the harmonic potential in Eq. 7.5. Such a potential induces a new nonlinearity in the dynamics, but the Langevin equation may again be relaxed by a mean-field approximation that incorporates the effects of the potential into the coefficients γ_p^Z and k_p^Z. This is not simple, however, because in a good solvent the static distribution of the polymer changes due to the repulsive self-interaction force and the attractive solvation force. The polymer swells in such conditions, and this affects the averages of the Oseen tensor and the repulsive potential.

Due to the fact that the Zimm model produces Langevin equations that are so similar to those of the Rouse model, we do not provide any more mathematical detail on these models. Later on when we compute FCS curves, we will again perform all of our derivations using the Rouse model. By inserting well-known alternative coefficients for the standard Rouse coefficients, we may apply our results to the Zimm model. Table 7.1 lists the proper coefficients for the Rouse model and for the Zimm model both in Θ conditions and in a good solvent. Their derivation can be found in [91].

7.1.5 Chain stiffness: semiflexible chain

A review of polymer theory in the context of the dynamics of double-stranded DNA would be incomplete without mention of the effect of chain stiffness on the polymer dynamics. No real

	Rouse	Zimm, Θ	Zimm, good
k_p	$6k_BT\pi^2p^2/L$	$6k_BT\pi^2p^2/L$	$k_BTb^{1/5}p^{11/5}/L^{6/5}$
γ_0	γL	$\frac{3}{8}\eta_s\sqrt{6\pi^3Lb}$	$\eta_sL^{3/5}b^{2/5}$
γ_p	$2\gamma L$	$\eta_s\sqrt{12\pi^3Lbp}$	$\eta_sb^{2/5}L^{3/5}p^{2/5}$

Table 7.1: Langevin equation coefficients for various flexible polymer models. η_s is the solvent viscosity. The Zimm/good solvent coefficients are all scaling relationships, hence the absence of numerical coefficients.

polymer chains behave like the flexible polymers we have described in this section over all size-scales. Rather, the bonds holding the chains together induce some stiffness that implies that on very small scales the Gaussian chain model breaks down: the vector between submolecules n and $n+1$ will necessarily resemble that between $n-1$ and n, because a lot of energy is needed to produce a sharply-kinked chain. More detailed models that incorporate chain-stiffness, such as the Kratky-Porod[91, 98] and semiflexible Gaussian chain model[99], are needed to accurately describe the polymer on these size scales.

The length scale over which the polymer stiffness is important is called the *persistence length*, and is a characteristic feature of specific polymer molecules. Winkler showed[23] that the ability of an FCS measurement to resolve chain stiffness depends on the waist of the probe beam being comparable to or larger than the persistence length. The persistence length of double-stranded DNA is about 50nm[100], so our probe beam waists of 400nm and 700nm are unlikely to produce FCS curves showing strong signatures of chain stiffness on the DNA molecules that we study. Furthermore, the extent to which semiflexible chain behavior dominates a polymer's motion depends on the relationship between the overall length of the polymer and the persistence length. This is easy to understand, since a polymer that is only as long as its persistence length is rather rodlike, with dynamics dominated by rotational modes. λ-phage DNA molecules are about 340 times longer than the persistence length, so that their dynamics are dominated by flexible chain-like motions (stretching modes) rather than stiff chain motions (bending modes, rotational modes). We do not discuss semiflexible chain models further in this thesis, but refer to [23] as a good review of the dynamics of these chains and particularly their application to FCS measurements.

7.2 Fluorescence correlation spectroscopy of polymers

For years, polymers have been studied by irradiating them with a probe beam and studying the scattered light. Regardless of the mechanism for this scattering — whether a neutron beam is scattered off of isotopes within the chain or a laser beam is absorbed and fluorescence emitted — the theoretical calculations for describing the scattering are almost identical. As defined in

Section 4.2.3, the *dynamic structure factor*

$$S(\mathbf{k},\tau) = \sum_{l,m} \left\langle e^{i\mathbf{k}^T(\mathbf{x}_m^{t+\tau}-\mathbf{x}_l^t)} \right\rangle, \tag{7.24}$$

where $\mathbf{x}_m^{t+\tau}$ and $\mathbf{x}_l^t$ are the positions of the l^{th} and m^{th} dye molecules (or scattering centers) is closely related to what is measured by FCS. We begin this section by presenting exact and approximate expressions for the dynamic structure factors of Rouse and Zimm polymers. Next we compute a related quantity, $\varphi(l,m;t,\tau)$ from Section 4.2.1, that is needed for tracking-FCS measurements.

7.2.1 Dynamic structure factor

Here we summarize the computation of $S(\mathbf{k},\tau)$ for the Rouse and Zimm models, closely following Doi and Edwards' review[91] of the calculations by de Gennes and Dubois-Violette [101, 102]. We derive an exact expression that applies to both models, and then perform a standard simplification using parameters from the Rouse model.

Since the statistics of the polymer are Gaussian, we may transfer the average in Eq. 7.24 from outside the exponential into the exponent according to Eq. 4.22. Doing this, while simultaneously accounting for the fact that there is no coupling between $\mathbf{R}(y,t)$ and $\mathbf{R}(y',t')$ along orthogonal axes, we get

$$S(\mathbf{k},\tau) = \frac{1}{N}\sum_{l,m} \exp\left[-\frac{1}{6}|\mathbf{k}|^2 \left\langle \left|\mathbf{x}_m^{t+\tau} - \mathbf{x}_l^t\right|^2 \right\rangle\right], \tag{7.25}$$

where the factor of $1/3$ comes from the fact that the inner product contains three quantities with identical statistics, one for each axis. It is useful shorthand (and convention[91]) to write

$$\phi(l,m;t,\tau) \equiv \left\langle \left|\mathbf{x}_m^{t+\tau} - \mathbf{x}_l^t\right|^2 \right\rangle, \tag{7.26}$$

which is related to $\varphi(l,m;t,\tau)$ as defined in Eq. 4.31, as we will see later in Section 7.2.3.

It is important to remember the distinction between $\mathbf{x}_m$, the position of the m^{th} dye, and $\mathbf{R}(y,t)$, the position of the polymer. In the dense-labeling limit, the positions of the dyes and of the polymer are indistinguishable; however, for more sparse labeling the continuous polymer chain and the discrete set of dyes are distinct and must be treated as such.

We may now insert Eq. 7.18 for the real-space Rouse chain dynamics into Eq. 7.25, which gives us

$$\phi(l,m;t,\tau) = \left\langle \left| \chi_0^{t+\tau} - \chi_0^t \right|^2 \right\rangle + 4 \sum_{p=1}^{\infty} \left\{ \left\langle \left(\chi_p^{t+\tau} \right)^T \chi_p^{t+\tau} \right\rangle \cos^2 \left[\frac{p\pi y_m}{L} \right] \right.$$
$$\left. + \left\langle \left(\chi_p^t \right)^T \chi_p^t \right\rangle \cos^2 \left[\frac{p\pi y_l}{L} \right] - 2 \left\langle \left(\chi_p^{t+\tau} \right)^T \chi_p^t \right\rangle \cos \left[\frac{p\pi y_m}{L} \right] \cos \left[\frac{p\pi y_l}{L} \right] \right\}, \quad (7.27)$$

where we have used the δ_{pq} term in Eq. 7.16 to preemptively eliminate all products between unlike Fourier coefficients. The $p = 0$ term in Eq. 7.27 corresponds to the center-of-mass motion given by Eq. 7.17. The remaining sum is simplified partly by the following peculiar fact[91]:

$$\sum_{p=1}^{\infty} \frac{1}{p^2} \left[\cos \left(\frac{p\pi y_m}{L} \right) - \cos \left(\frac{p\pi y_l}{L} \right) \right]^2 = \frac{\pi^2}{2L} \left| y_m - y_l \right|. \quad (7.28)$$

After inserting this and the covariance for χ_p from Eq. 7.16, we are left with

$$\phi(l,m;t,\tau) = 6D_g\tau + \left| y_m - y_l \right| b + 24k_BTb \sum_{p=1}^{\infty} \frac{1}{k_p} \left(1 - e^{-\tau/\tau_p} \right) \cos \left[\frac{p\pi y_m}{L} \right] \cos \left[\frac{p\pi y_l}{L} \right], \quad (7.29)$$

where τ_p is defined in Eq. 7.22. When calculated using Eq. 7.29, the dynamic structure factor is exact for the Rouse and Zimm models. However, it is also not easy to calculate, requiring numerical evaluation of the sums over l, m and p. If we study Eq. 7.29 in certain limits, we may evaluate its approximate behavior.

In the so-called small-angle regime defined by $|\mathbf{k}|^2 Nb^2 \ll 1$, the only term that is not very small in Eq. 7.29 is the first: we get $S(\mathbf{k},\tau) = Ne^{-|\mathbf{k}|^2 D_g \tau}$. This is not at all surprising, since it just implies that the low-frequency modes are dominated by Brownian motion of the center of mass of the polymer.

The most common approximation used in fluorescence measurements of polymers [21, 25] is that the measurement time-scale is much shorter than the polymer chain's relaxation time-scale, so $\tau \ll \tau_1$. Expanding the exponential term inside the sum over p then allows us to show that the sum is dominated by large p, allowing a simplification of the cos terms. We write the identity

$$2\cos \left[\frac{p\pi y_m}{L} \right] \cos \left[\frac{p\pi y_l}{L} \right] = \cos \left[\frac{p\pi (y_m - y_l)}{L} \right] + \cos \left[\frac{p\pi (y_m + y_l)}{L} \right], \quad (7.30)$$

where the second term fluctuates very rapidly at the large p values that contribute the most to the sum; for this reason we ignore it, assuming that its fluctuations will average away to zero. The resulting expression for $S(\mathbf{k},\tau)$ is written in terms of three discrete sums. We can simplify it by converting those sums into integrals. From here on, we use the values for k_p and τ_p for the Rouse model, but the derivation for the Zimm model is done similarly. For more details, see the original derivation[102].

For the Rouse model we have

$$S(\mathbf{k},\tau) \approx \frac{1}{L}\int_0^L \mathrm{d}y_m \int_{-y_m}^{L-y_m} \mathrm{d}u \exp\left\{-|\mathbf{k}|^2 D_g \tau - \frac{1}{6}|\mathbf{k}|^2|u|b \right. \\ \left. - \frac{Lb}{3\pi^2}|\mathbf{k}|^2 \int_0^\infty \mathrm{d}p \frac{1}{p^2} \cos\left[\frac{p\pi u}{L}\right]\left(1 - e^{-\tau p^2/\tau_1}\right)\right\}, \quad (7.31)$$

where we have used $u = y_l - y_m$. From here, Doi and Edwards make the argument that the integrand is sharply peaked at $u = 0$, so that the integral over u only depends on y_m when y_m is near zero. We follow their work here, but we will find in the next section that this introduces a large error in $S(\mathbf{k},\tau)$ for small $\mathbf{k}$. This is significant for us, because the FCS curve is dominated by the small-$\mathbf{k}$ terms. If indeed the integrand is sharply peaked, then for large enough L the y_m dependence of the integration bounds contributes very little relative to the overall size of the integral. We may compute the integral over u across the entire real line without drastically affecting the result. We make some variable substitutions to get

$$S(\mathbf{k},\tau) = \frac{12}{|\mathbf{k}|^2 b} e^{-|\mathbf{k}|^2 D_g \tau} \int_0^\infty \mathrm{d}v \exp\left\{-v - (\Gamma_k \tau)^{1/2} h\left[v\,(\Gamma_k \tau)^{-1/2}\right]\right\}, \quad (7.32)$$

where we have defined

$$\Gamma_k \equiv \frac{k_B T}{12\gamma} |\mathbf{k}|^4 b^2 \quad (7.33)$$

and

$$h(v) \equiv \frac{2}{\pi} \int_0^\infty \mathrm{d}x \, \frac{\cos(xv)}{x^2} \left(1 - e^{-x^2}\right). \quad (7.34)$$

A final approximation is now used to convert Eq. 7.32 into a more useful form. If we assume that $\Gamma_k \tau \gg 1$, then we may substitute $h(0)$ for the term inside the exponential, in which case we may compute the integral:

$$S(\mathbf{k},\tau) \approx \frac{12}{|\mathbf{k}|^2 b^2} \exp\left[-|\mathbf{k}|^2 D_g \tau - \frac{2}{\sqrt{\pi}} (\Gamma_k \tau)^{1/2}\right]. \quad (7.35)$$

This approximation receives widespread use in polymer dynamics studies. It can be shown that the short-time approximation for the dynamic structure factor for the Zimm polymer in a good solvent yields a similar expression, but with an exponent with a $\tau^{2/3}$ dependence. By fitting the temporal decay of the measured dynamic structure factor to a stretched exponential of the form $S(\mathbf{k},\tau) \propto \exp\left[-|\mathbf{k}|^2 (\Gamma \tau)^\beta\right]$, several groups claim to have found evidence of Zimm or Rouse behavior in their measurements[21, 22, 25].

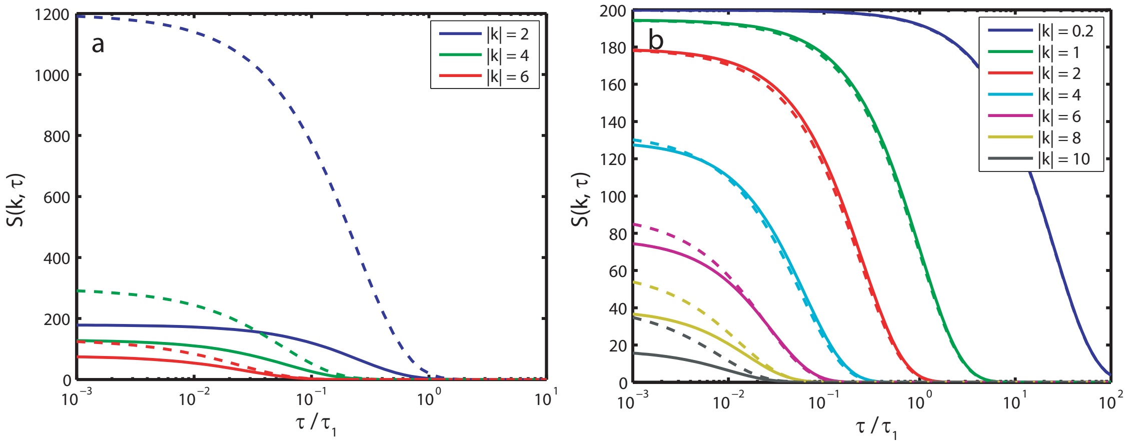


Figure 7.4: Exact (solid) and approximate (dashed) Rouse dynamic structure factor. **a** compares the exact $S(\mathbf{k},\tau)$, computed numerically in Matlab, with the approximation from Eq. 7.35. **b** compares the exact $S(\mathbf{k},\tau)$ to our corrected approximation from Eq. 7.37. The approximate curve for $|\mathbf{k}| = 0.2$ is not clearly visible because its deviation from the exact curve is very small. Numerical computations truncated the sum over p at $p = 200$. We used $b = 0.05$, $L = 10$, $D_g = 1$ [arbitrary units] in all calculated curves.

7.2.2 Application to open-loop FCS

Equation 7.35 was derived through the use of a large number of approximations. This standard derivation was well-suited to dynamic light scattering experiments, but we must consider their effect on FCS measurements. In particular, because light scattering measurements are made at a fixed **k** vector and FCS measurements are an integral over all **k** vectors, there may be some difficulty in directly using Eq. 7.35 to compute our FCS curves.

The prefactor of $|\mathbf{k}|^{-2}$ in Eq. 7.35 is very significant. Because of it, if we were to attempt to compute the FCS curve according to Eq. 4.41, the integral would not converge. In fact, the error at $\tau = 0$ for small **k** vectors is enormous, as shown in Fig. 7.4**a**. This error is not a problem at all for DLS measurements at a fixed **k** vector, because the normalization of $S(\mathbf{k}, 0)$ is found experimentally.

We may backtrack a few steps and find an expression for $S(\mathbf{k},\tau)$ that is more acceptable for use in FCS. If we had not eliminated the dependence on y_m on the integration limits in going from Eq. 7.31 to Eq. 7.32, but we had made the assumption that $\Gamma_k\tau \gg 1$, then we would have

$$S(\mathbf{k},\tau) \approx \frac{1}{L}\int_0^L \mathrm{d}y_m \int_{-y_m}^{L-y_m} \mathrm{d}u \exp\left[-|\mathbf{k}|^2 D_g\tau - \frac{1}{6}|\mathbf{k}|^2|u|b - \frac{2}{\sqrt{\pi}}(\Gamma_k\tau)^{1/2}\right] \tag{7.36}$$

$$= \frac{12}{|\mathbf{k}|^2 b^2}\left[1 - \frac{6}{N|\mathbf{k}|^2 b^2}\left(1 - e^{-|\mathbf{k}|^2 b^2 N/6}\right)\right]\exp\left[-|\mathbf{k}|^2 D_g\tau - \frac{2}{\sqrt{\pi}}(\Gamma_k\tau)^{1/2}\right] \tag{7.37}$$

for the Rouse model. The results of this correction are shown in Fig. 7.4**b**. The corrected curve is a very good approximation to the exact $S(\mathbf{k},\tau)$, especially for the small $|\mathbf{k}|$ terms that contribute the most to the FCS curve. Errors at $\tau = 0$ reappear at larger $\mathbf{k}$ because of our approximation of the fast-decaying sums by integrals in Eq. 7.31. Fortunately, the $|\mathbf{k}|^{-2}$ decay in the amplitude of $S(\mathbf{k}, \tau)$ means that these less-accurate large-$\mathbf{k}$ terms contribute only a small amount to the FCS curve. The agreement between the approximate and exact curves in the figure is striking at times comparable to τ_1, since we obtained Eq. 7.37 via the short-time approximation. If we plot the relative error between the exact and approximate curves at a fixed $|\mathbf{k}|$ (not shown) we find that the error does increase on longer time-scales, but only after $S(\mathbf{k}, \tau)$ has almost completely decayed to zero. This means that any fears concerning the use of an approximation for $S(\mathbf{k}, \tau)$ for measurements spanning long time-scales are unwarranted for the Rouse polymer. The Zimm polymer in good solvent is more difficult to analyze in this way because of the approximate nature of the coefficients in Table 7.1.

We now compute the the integral in Eq. 4.41 to find the FCS curve. If we insert a stationary Gaussian laser beam, as described by Eq. 4.11, we find that in general a closed-form expression for the integral does not exist. However, if the beam waist w/ζ along the z axis is comparable to w, the waist along the x and y axes, then we may compute the integral using a transformation to spherical coordinates. This gives

$$\int \frac{\mathrm{d}^3\mathbf{k}}{(2\pi)^3} S(\mathbf{k}, \tau)\tilde{\Gamma}(\mathbf{k}) \approx \frac{\Gamma_0 w^3 \pi^{3/2}}{32 b^2} \frac{1}{\sqrt{\nu(\tau)}} \left\{1 + 2\lambda\nu(\tau) - 2\sqrt{\lambda\nu(\tau)\left[1 + \lambda\nu(\tau)\right]}\right\}, \tag{7.38}$$

where we have defined $\lambda = 6/Nb^2$ and

$$\nu(\tau) = D_g\tau + \frac{2}{\sqrt{\pi}}\left(\frac{k_B T b^2}{12\gamma}\tau\right)^{1/2} + \frac{w^2}{8}. \tag{7.39}$$

This calculation gives us a fair estimate of what the FCS curve will look like, but the requirement that $\zeta = 1$ has its consequences: the decay in the predicted $g_2(\tau)$ will be steeper, and the predicted contrast will be greater, than the measured curve. More generally, the approach used in Chapter 4 and in [23, 103] should be used: the integrals over $\mathbf{k}$ are computed first to find the FCS curve in terms of $\phi(l, m; t, \tau)$, and then the polymer physics is inserted and the FCS curve itself is approximated as needed.

7.2.3 Tracking-FCS of polymers

So far in this section we have only calculated curves that are useful for other people; we showed in Section 4.2.3 that the dynamic structure factor does not apply directly to our tracking-FCS measurements. In fact, we showed that the only tracking measurements for which it does apply are camera-based correlation methods where the illumination beam intensity is uniform over the region explored by the tracked particle.

From Eq. 4.31, we know that the tracking-FCS curve is determined by the quantity

$$\varphi(l,m;t,\tau) = \left\langle \mathbf{x}_m^{t+\tau} \left(\mathbf{x}_l^t\right)^T \right\rangle . \tag{7.40}$$

In both of our models this outer product matrix is diagonal and the motion of the chain is isotropic; therefore we know that

$$\varphi(l,m;t,\tau) = \frac{1}{3} \left\langle \left(\mathbf{x}_m^{t+\tau}\right)^T \mathbf{x}_l^t \right\rangle . \tag{7.41}$$

In our derivation of the tracking-FCS curve in Section 4.2.1 we defined our coordinate system so that $\left\langle \mathbf{x}_l^t \right\rangle = 0$. Using this and the statistics we calculated for chain dynamics in Section 7.1.3, we may represent $\varphi(l,m;t,\tau)$ exactly by the series

$$\varphi(l,m;t,\tau) = 4k_BTb \sum_{p=1}^{\infty} \frac{1}{k_p} \exp\left[-\frac{k_p}{\gamma_p}\right] . \tag{7.42}$$

In order to compute the tracking-FCS curve, we require what we referred to as the radius of gyration R_g of individual scattering centers along the backbone of the polymer chain. These are defined by

$$R_g(m)^2 \equiv \left\langle \left| R(y_m,t) - R_{cm}(y_m,t) \right|^2 \right\rangle , \tag{7.43}$$

and so in the standard way we insert the statistics of the polymer in terms of k_p:

$$R_g(m)^2 = 4k_BTb \sum_{p=1}^{\infty} \frac{1}{k_p} \cos^2\left[\frac{p\pi y_m}{L}\right] . \tag{7.44}$$

In the case of either the Rouse polymer or the Zimm polymer in Θ conditions, this series is another peculiar one with a simple closed-form solution. I found this expression by trial and error, although I have no doubt that it has been known for a long time:

$$R_g(m)^2 = \frac{Lb}{3}\left[\left(\frac{y_m}{L} - \frac{1}{2}\right)^2 + \frac{1}{12}\right] . \tag{7.45}$$

Especially due to the strong agreement between the exact and approximate dynamic structure factors for the Rouse polymer as illustrated in Fig. 7.4, it is tempting to try to relate $\varphi(l,m;t,\tau)$ to $\phi(l,m;t,\tau)$, because of the convenient approximations for $\phi(l,m;t,\tau)$. If we expand the squared term in Eq. 7.26 and isolate the cross-term, we get

$$\varphi(l,m;t,\tau) = \frac{1}{6}\left[R_g(m) + 6D_g\tau + R_g(l) - \phi(l,m;t,\tau)\right] \mathrm{Id}_3 . \tag{7.46}$$

Now it seems appropriate to insert

$$\phi(l, m; t, \tau) = 6D_g\tau + |y_m - y_l|b + \frac{12}{\sqrt{\pi}}\left(\frac{k_B T b^2}{12\gamma}\tau\right)^{1/2} \tag{7.47}$$

into Eq. 7.46; however, this ϕ increases without bound as $\tau \to \infty$, and so it will clearly produce invalid results: Eq. 4.33 predicts that the correlation function would approach -1 at long times because the normalization constant cannot depend on τ. This approximation clearly produces invalid tracking-FCS curves.

7.3 Literature review

Now, in the context of the theory that we have reviewed in this chapter, we review the recent literature that is relevant to our measurements.

7.3.1 Lumma et al., *Phys. Rev. Lett.*, 2003

This paper[21] was the first to use FCS to study the internal dynamics of DNA molecules in solution. The authors studied λ-phage DNA labeled along its backbone by intercalating dyes. They postulated that intramolecular dynamics would influence the FCS curve because their focused laser beam waist ($\sim$ 180nm) was much smaller than the radius of gyration of the polymer (730nm). Indeed, their FCS curves decayed on shorter time-scales than would be predicted for a solid particle of this size and decayed with a functional form that differed slightly from that which they predicted for pure Brownian motion. They fit curves based on a theory for a semiflexible polymers and a short-time approximation for the dynamic structure factor. Their measurements showed evidence of chain stiffness influencing the FCS curves in the form of an inferred dynamic structure factor scaling as $\exp\left[-|\mathbf{k}|^2 (\Gamma\tau)^{3/4}\right]$ (characteristic of a semiflexible chain) when the average distance between dyes was comparable to the persistence length of the polymer.

I have three concerns about the results in this paper. First, open-loop FCS was used and so translational diffusion contributed to the decay of the FCS curve. The polymer's internal dynamics, it was argued, decay on shorter time-scales than the translational motion; such time-scale separation allows both processes to be resolved simultaneously. However, the fits to the data yielded translational diffusion coefficient estimates ($(1.01 \pm 0.05)\mu\text{m}^2/\text{s}$) that were much larger than other measurements have produced. While this may have been the result of an inaccurate calibration of the FCS setup, in which case the decay characteristics of $g_2(\tau)$ would not be affected, it is not clear that the problem was not more fundamental. It is difficult to accurately extract dynamics from FCS curves when those dynamics overlap in time.

Second, the theoretical fits to the data in this paper were derived using the expression for the dynamic structure factor $S(\mathbf{k}, \tau) = \exp\left[-|\mathbf{k}|^2 (\Gamma\tau)^\beta\right]$. This approximation is appropriate only on short time-scales, but was used quite generally without such limitation; this may lead to misleading results, as discussed in the next section and in [24].

Finally, as was shown in Section 7.2.2, the **k**-dependent prefactor of the dynamic structure factor must be considered in computing the FCS curve. Due to the most common use of the dynamic structure factor in measurements at a fixed **k**, the prefactor is sometimes omitted; however, when integrating over **k**, it should not be left out. Exactly how severe the consequences of this omission are is not clear. Detailed comparison of the $g_2(\tau)$ derived using the exact expressions for $S(\mathbf{k}, \tau)$ are necessary to determine this.

7.3.2 Shusterman et al., *Phys. Rev. Lett.*, 2004 and Petrov, et al., *Phys. Rev. Lett.*, 2006

The first of these papers[22] studied both single- and double-stranded DNA molecules using FCS by labeling them at their ends with a single fluorescent dye. The authors analyzed the FCS curves using the short-time approximation to the dynamic structure factor, finding stretching exponents that uniformly predict Rouse behavior ($\beta = 1/2$) for double-stranded molecules and Zimm behavior ($\beta = 2/3$) for single-stranded molecules. They describe these results as puzzling, and rightly so: hydrodynamic interactions are expected to increase with polymer length for semiflexible polymers because more interaction takes place between distant parts of the chain. They observe the opposite, inferring from their data that Rouse dynamics dominate even more strongly on longer chains.

The second of these papers[24] addresses concerns that those authors had with the previous paper. They perform similar measurements, FCS on end-labeled double-stranded DNA molecules of different sizes, but perform their analysis quite differently and come to quite different conclusions: they find evidence of strong hydrodynamic interactions and that the molecules they study are semiflexible chains. The authors of this paper argue, correctly, that the authors of the first paper misapply the predictions of the short-time approximation over long time-scales. By contrast, the authors of the second paper numerically calculate the exact form of the FCS curve and fit it to their data; their results fit the theory for semiflexible polymers quite well.

7.3.3 Cohen and Moerner, *Phys. Rev. Lett.*, 2007

This paper[25] distinguishes itself from the others we have discussed in that feedback control was used to eliminate contributions due to translational diffusion. λ-phage DNA molecules labeled with intercalating dyes and studied by fluorescence video microscopy while they were immobilized in an electroosmotic trap. The video images were analyzed with an autocorrelation method as described in Section 4.2.3, as well as with spatially-resolved measurements that directly probe the dynamic structure factor of the molecules. Curves fit to both measurements led to the conclusion that Zimm-type hydrodynamic interactions determine the dynamics of the polymer; the spatially-resolved measurements implied that Zimm behavior dominated the higher-frequency spatial modes and that Rouse behavior dominated the lower-frequency modes. These were the first measurements to suggest such $\mathbf{k}$ dependence of the qualitative features of the polymer dynamics, first predicted by Dubois-Violette and de Gennes[102]. Additionally, these measurements showed an interesting radial dependence on the relaxation of the polymer's dynamics, and are the only measurements to show correlation between a polymer's instantaneous radius and its time-dependent diffusion coefficient.

I have several concerns about this paper. First, there are several flaws in the analysis of the image-image correlation function $C(\tau)$. As we discussed in Section 4.2.3, the application of the dynamic structure factor to the image-image correlation function $C(\tau)$ only applies if the excitation beam's radius is very large compared to that of the molecule under study. Furthermore, the point spread function of the microscope factors into $C(\tau)$ but was omitted from the analysis without justification. Finally, the curve that was fit to $C(\tau)$ in this paper is either incorrect, or the paper contains a typographical error. A curve of the form $S(\tau) = \exp\left[-(\gamma\tau)^{\beta}\right]$ was fit to $C(\tau)$, and conclusions were drawn by comparing the fit value β to the values (1/2, 2/3) predicted for the exponents of the short-time approximate dynamic structure factors. Assuming that the excitation beam was large enough, the correct fit function should have been the integral of the product of the dynamic structure factor and point spread function, as discussed in Section 4.2.3. This yields the substantially different scaling $C(\tau) \sim \left[1 + (\gamma\tau)^{3/2}\right]^{-3/2}$.

Second, the analysis of the dynamic structure factor as measured from the camera images may be incorrect. The measurements in this paper are intrinsically long time-scale measurements; fluorescence correlation is resolved over time-scales approaching 1s. However, the short-time approximation was used for all theoretical fits. As argued in [24], the indiscriminate application of these approximate expressions for the dynamic structure factor can provide misleading results; the conflicting results of [22, 24] illustrate this quite clearly. Over the extremely long time-scales of a tracking/trapping measurement, it is very difficult to justify the application of these approximations.

Third, tracking error statistics are not accounted for in these measurements. This is troubling because it is claimed that the apparatus was only able to trap each molecule for between 9 and 18 seconds. There is theoretically no limit to the duration for which a particle may be trapped in an electroosmotic trap, unlike the limits imposed on our apparatus by the finite travel of the tracking stages. The fact that these molecules were lost implies that either they were always tracked very poorly and escaped (see Section 3.3.2) or, more likely, the tracking fidelity decayed over time due to photobleaching of the molecules. Either way, tracking errors should have been accounted for. This is especially true since the measurements were spatially resolved: tracking errors in a uniform illumination profile are invisible to FCS, but easily seen by a camera.

Finally, the two-dimensional nature of the trap raises questions about the validity of the measurements due to interactions with the cell walls. The λ-phage DNA molecules extend over a large fraction of the trap and are in almost constant contact with the trap walls. This is reflected by the diffusion coefficient estimate ($(0.32 \pm 0.02)\,\mu\text{m}^2/\text{s}$) that is much smaller than other measured values. Without comparisons to measurements made on freely-diffusing molecules, it cannot be known how these interactions affect the dynamics of the molecules.

7.4 Measurements

We purchased λ-phage DNA (48514bp, $L \sim 17\mu$m) from New England BioLabs (Ipswich, MA) and diluted it to a concentration of about 160pM in a buffer containing 10mM TRIS, 10mM NaCl and 1mM EDTA in 18MΩ purified water at pH 8.0. We purchased the intercalating dyes POPO-1 (434nm/456nm) and POPO-3 (534nm/570nm) from Invitrogen/Molecular Probes (Eugene, OR). We diluted the POPO-3 dyes to 100pM and the POPO-1 dyes to a concentration between 2.5pM and 1nM in TRIS buffer with 140mM 2-mercaptoethanol added to improve photostability. We prepared samples with a final concentration of DNA molecules of about 300fM by diluting the DNA further into the dye solutions. This corresponded to dye label densities ranging between 1 dye : 15bp and 1 dye : 6000bp. We incubated the dye/DNA solutions for 15 minutes in the dark at room temperature to ensure complete attachment; longer incubation times did not produce noticeably different brightnesses among labeled molecules using the POPO dyes.

We tracked a total of 59 different molecules for times ranging between 5 and 54 s. We used the fluorescence servo described in Section 5.2.6 to modulate the power in the 532nm beam and keep the measured fluorescence rate of the POPO-3 dyes fixed at about 10^4 photons/sec. Resulting beam powers reached as low as about 20nW, and were limited to about 10μW using neutral density filters. By fixing all of the systematic contributions to the fluorescence data on this detection channel, we ensured that the tracking error statistics were the same for each molecule that we tracked.

We set the optical power in the violet probe beam to a fixed value for each experiment using a combination of neutral density filters and the probe intensity stabilization servo described in Section 5.2.1. We tried to set the power to a level that maintained an average fluorescence

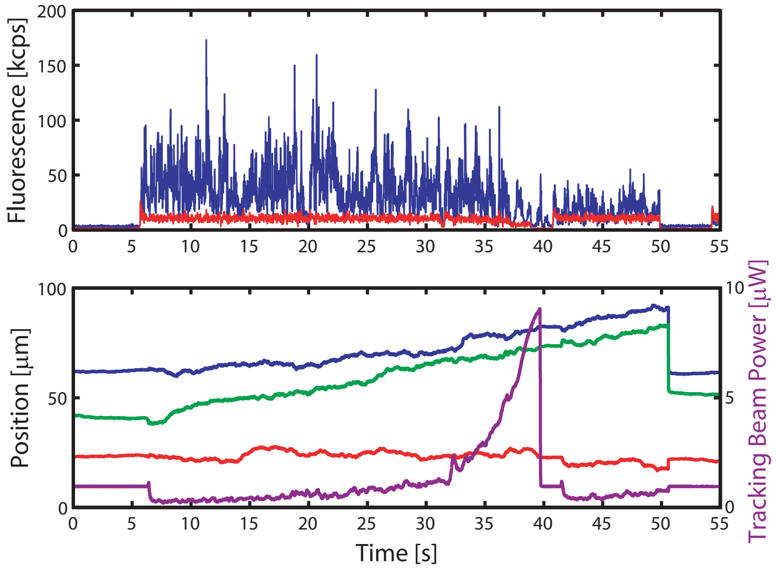
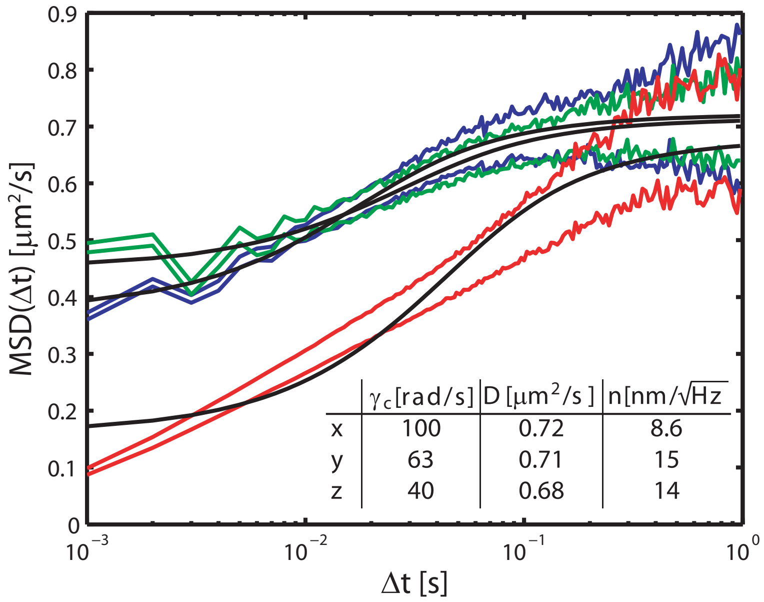


Figure 7.5: Example tracking data with a label density of 1 dye : 300bp and 415nm illumination power of ~ 800nW. **Top**: fluorescence intensity in the blue and red fluorescence channels. **Bottom**: x (blue), y (green) and z (red) tracking stage positions over time. The violet curve shows the optical power of the 532nm beam, which was controlled by the fluorescence servo.

rate between 10^4 and 10^5 photons/sec: by collecting fluorescence at this rate for ~ 10s, we achieve high statistical accuracy in our FCS curves on time-scales longer than ~ 10μs. The optical power needed was determined by the label density, ranging between about 200nW and 2μW. At these powers, we see no evidence of triple state blinking or photobleaching of the blue dyes.

Figure 7.5 shows a typical data acquisition. The sharp contrast in the qualitative appearance of the blue and red fluorescence signals arises from the differences in the focused beam waists of the 415nm beam and 532nm beams. The 415nm beam is significantly smaller than the radius of gyration of the molecule, so individual dyes diffuse into and out of the beam due to the polymer's internal motion and produce large spikes in fluorescence. By contrast, the 532nm beams are significantly larger than the polymer's radius of gyration so the internal motion of the polymer has a much smaller affect on the fluorescence fluctuations.

The data in Fig. 7.5 clearly illustrates the functioning of the fluorescence lock servo. As shown in the lower plot, the fluorescence rate remains locked at a value well below its maximum while no particle is being tracked ($t < 5$s). Once a molecule drifts into the focused laser, its fluorescence causes a spike on the red channel. The servo responds by reducing the beam power and continues adjusting the power over the course of the data run, increasing it in response to the bleaching of the POPO-3 dyes. Between $t = 35$s and $t = 40$s, the apparatus loses the DNA molecule because the fluorescence servo is incapable of further increases in optical power. Once the fluorescence rate drops, the servo quickly reduces the optical power to its set-point



	γ_c[rad/s]	D [μm²/s]	n[nm/√Hz
x	100	0.72	8.6
y	63	0.71	15
z	40	0.68	14

Figure 7.6: Average interval-dependent mean-squared deviation of the tracking stages for 42 DNA molecules, representing a total of 859s of tracking data. We plotted the 2σ error bounds of the data to illustrate the experimental uncertainty; for clarity, only the bounds (not the mean values) are shown. We fit curves to the data using Eq. 3.35, with $\gamma_p = \infty$ for all curves because we do not see obvious evidence of the plant dynamics so the second-order model is too rich for the data. The inset table provides the remaining fit parameters.

in the absence of tracking; a second DNA molecule quickly drifts into focus, but is only tracked for a few seconds because the x stage reaches its upper boundary and all stages reset. The difference in the brightness of the blue fluorescence from the two molecules suggests that the first molecule was labeled more densely than the second. We are not certain why we observe such variations in label density within a single sample.

The distinctly nonexponential increase in the power required to lock the fluorescence rate arises for the following reason: the number of active dyes decays exponentially with a rate that is proportional to the excitation power. As the excitation power increases to track the decaying fluorescence, the decay rate increases. We expect the excitation power $P(t)$ to obey the nonlinear differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}P(t) = \Gamma_0 - P(t)\exp\left[\int_0^t \mathrm{d}t' P(t')\right] \tag{7.48}$$

which, to my knowledge, does not have a clean closed-form solution. Exact quantitative understanding of $P(t)$ is not so important, so we will not analyze this any further.

7.4.1 Center of mass motion

First we analyze the center of mass motion of the DNA molecules that we tracked by considering the motion of the tracking stages. Figure 7.6 shows $MSD(\Delta t)$ as defined by Eq. 3.33 along with theoretical fits to Eq. 3.35. We averaged the data from 42 separate molecules, with tracking times ranging between 7 and 54s. Some data had to be excluded because the z tracking stage had railed on its upper boundary; in these cases, the data along all three axes were rejected in order to avoid biasing the results with data from molecules that were tracked away from the localization origin.

Our fits to the $MSD(\Delta t)$ curves give $D = (0.71 \pm 0.05)\,\mu\text{m}^2/\text{s}$, corresponding to a polymer with a radius of gyration of about 600nm. The dominant source of error in our fits is not fitting error, but statistical uncertainty in computing $MSD(\Delta t)$. The uncertainty $0.05\mu\text{m}^2/\text{s}$ that we specify is the mean of the 2σ bounds shown about the measured curves at $\Delta t = 1\text{s}$. It is easy to calculate what we expect the uncertainty in $MSD(\Delta t)$ to be for a finite number of measurement intervals. We calculate $\langle MSD(\Delta t)^2\rangle = 3D_g^2$ for pure Brownian motion; at times Δt that are long enough for our stage positions to be close approximations of the particle's actual position, this expression holds for the stage position statistics. The 2σ error bounds that we predict are therefore given by $\pm 2D_g\sqrt{3/N}$, which for our mean diffusion coefficient and 859 measurement intervals, is $\pm 0.08\mu\text{m}^2/\text{s}$. This is consistent with our measurements, implying that the molecules we tracked all had the same diffusion coefficient. If, for example, the DNA sample had been contaminated by a nuclease that digested the molecules into fragments of different lengths, we would have observed a larger uncertainty in the estimated diffusion coefficient.

While the fits to $MSD(\Delta t)$ are very good for both the x and y axes, the z-axis fits are clearly imperfect. The error in the fits is not simply one of poor fitting: we have observed similar z-axis curves in other measurements that the theory does not fit to. More investigation is needed to determine the source of this discrepancy. Fortunately, the z axis diffusion coefficient estimate cannot be far off because long-time features of $MSD(\Delta t)$ are not affected by the stage dynamics.

Our measured value for D_g is incompatible with all of the previously-measured diffusion coefficients for the λ-phage DNA molecule. In particular, dynamic light scattering yields $(0.41 \pm 0.05)\mu\text{m}^2/\text{s}$; wide-field fluorescence microscopy yields $(0.47\pm0.03)\mu\text{m}^2/\text{s}$ (or $(0.59\pm0.04)\mu\text{m}^2/\text{s}$, if scaled by $1.75^{2/5}$ as suggested by the authors to approximate that for unlabeled molecules)[104]; and open-loop FCS yields $1.1\mu\text{m}^2/\text{s}$[23] or $(1.01\pm0.05)\mu\text{m}^2/\text{s}$[21]. Closed-loop tracking experiments are the best way to measure diffusion coefficients, particularly for particles with internal dynamics, because the measured stage positions are completely decoupled from the particles' internal motion. FCS and DLS measurements only yield a convolution of the center of mass and

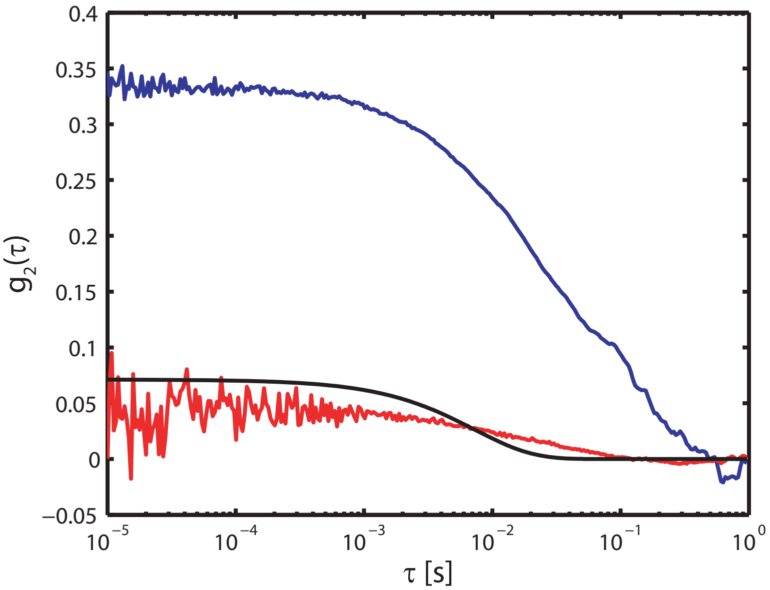


Figure 7.7: Example tracking-FCS curves, for the data in Fig. 7.5. The black curve is the tracking error contribution to the blue curve predicted by the theoretical fit parameters from Fig. 7.6.

internal motions. Accurate deconvolution relies on a strong understanding of both motions or, ideally, a large time-scale separation between them. Furthermore, FCS measurements are very sensitive to calibration, because it is impossible to know the exact focused beam intensity inside the sample. The only measurements similar to ours that have produced D_g values for λ-phage were those by made in a two-dimensional electroosmotic trap[25]. These measurements were affected by the shallow depth of the trap, reducing the diffusion coefficient by 24% from the smallest previously-measured value. In addition, those measurements relied on reconstructed particle trajectories based on applied velocity pulses, and it is not clear how accurately such reconstructions may be calibrated. By contrast, our nanopositioner stages can be calibrated externally with high-accuracy methods such as laser interferometry. We believe that our measurements are therefore the most precise yet made on the diffusion coefficient of this molecule.

7.4.2 Internal motion

Next we use tracking-FCS to study the internal motion of the DNA molecules using the same data described in the previous section. Our first consideration must be the effect of tracking errors on the FCS curves. The RMS tracking error is found from Eq. 3.31 by inserting the system parameters extracted from the $MSD(\Delta t)$ curves in the previous section: $\sigma_x^0 = 100$nm, $\sigma_y^0 = 140$nm, $\sigma_z^0 = 150$nm. These errors are small, but nontrivial compared to the focused 415nm beam waist: for a waist of 400nm and a polymer with radius of gyration of 600nm, the

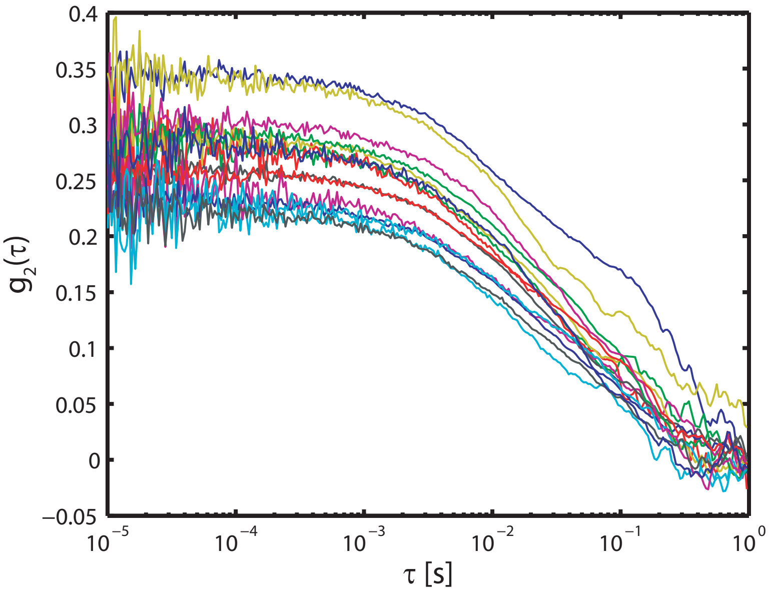


Figure 7.8: Variation in measured $g_2(\tau)$ curves with label density of 1 dye : 150bp.

2σ intensity fluctuations may be as large as about 10% of the peak fluorescence intensity.

Figure 7.7 shows the FCS curves for the first molecule tracked in Fig. 7.5 for both fluorescence channels, as well as the contribution to the blue curve due to the tracking error that we calculate from the fits in Fig. 7.6. The large contrast in the blue curve and different decay time-scales — the error decays at the feedback system bandwidth while the fluorescence decays at the polymer decay time — implies that we do see dynamics other than tracking errors in our signals. We note that the tracking error contribution to the red curve is nearly zero due to the much larger focused waist of the 532nm beam. The decay in the curves for both channels therefore represent the same statistics, with the exception of the fact that the red curve is influenced by the fluorescence servo, and so highlights the improvement in contrast due to our use of the tightly-focused 415nm beam. The curves appear to decay with similar time-scales, supporting this argument. The dominant decay has a time constant of about 100ms, which is comparable to that which we predict (180ms) for a freely-diffusing particle with its diffusion coefficient and radius of gyration but without internal dynamics according to the open-loop equivalent of the derivation in Section 4.2.1.1.

As illustrated in Fig. 7.6 and discussed in the last section, some of the molecules that we tracked were apparently less densely labeled than others from an identical sample. Accordingly, we find variations in the FCS curves. Figure 7.8 shows the FCS curves for the 15 molecules we tracked with a label density of 1 dye : 150bp. As reported in [23], decreased label density affects the FCS curve for flexible polymers by shifting it to the left. In our case, we see evidence that the curves are shifted upward. We attribute this to differences in the signal to background ratios

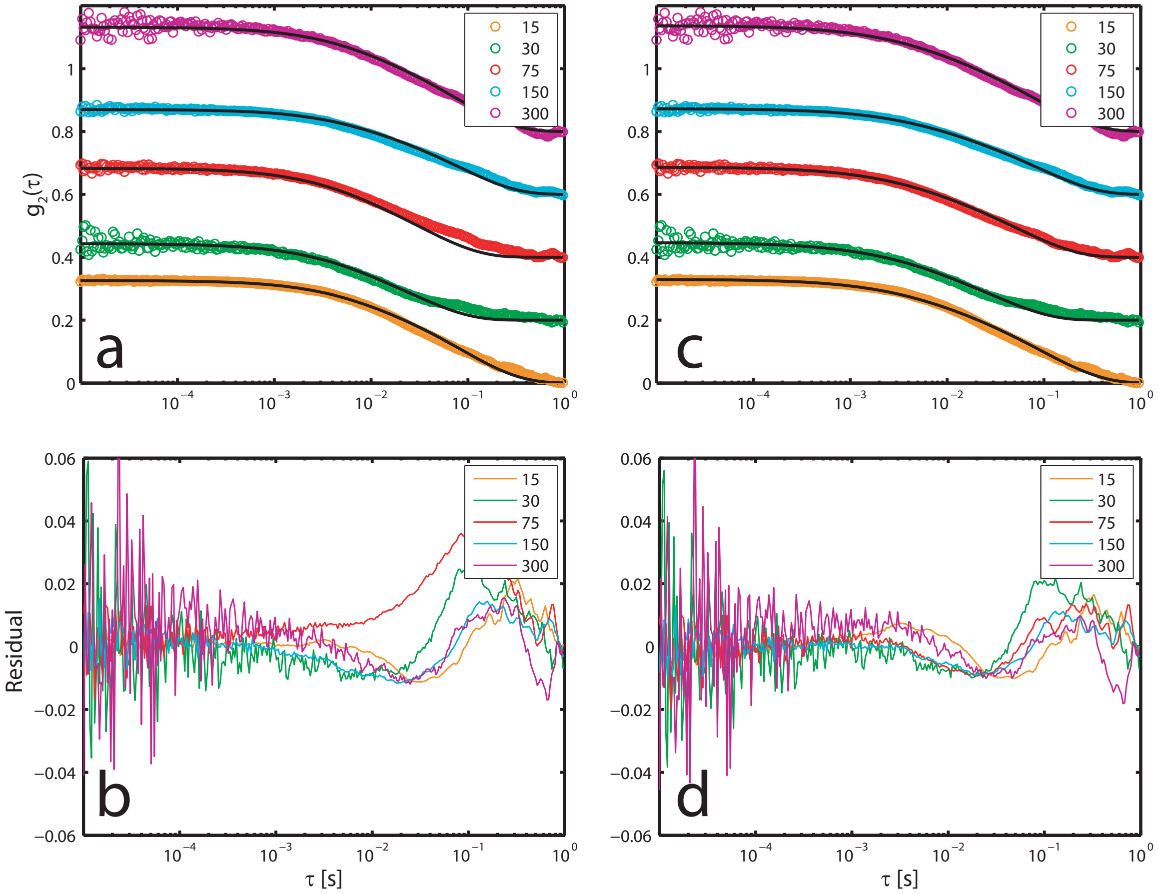


Figure 7.9: Fits (black) and residuals of Rouse and Zimm theories to measured $g_2(\tau)$ curves (colored circles). **a**: Fits to Zimm theory. **b**: Residuals of Zimm fit. **c**: Fits to Rouse theory. **d**: Residuals of Rouse fit. $g_2(\tau)$ plots offset successive curves by 0.2 for clarity. Label density (in bp/dye) is given in the legends.

of the differently labeled molecules. As we discussed in Section 4.1.4, increased background relative to signal causes an overall attenuation of the FCS curve.

We fit curves to our data for five different dye label densities using the exact expressions in Eq. 7.44 for $R_g(m)$, Eq. 7.42 for φ, and Eq. 4.33 for $g_2(\tau)$. We allowed for two adjustable parameters in the fits: b and τ_1 were set independently by the fitting routines. We fit to both the Rouse model and the Zimm model for a good solvent and found that both models produced comparable fits that are fairly good. Neither model appears to perfectly reproduce the decay on very long time-scales (near 1s), at which the measured curves turn upward and the theoretical curves do not. This feature of the experimental curves could be an artifact due to a small amount of photobleaching of the blue dyes; photobleaching would cause an increased contrast in $g_2(\tau)$ that decayed on the time-scale of the bleaching process.

Our results in this section leave us without an obvious conclusion. It is possible that the Zimm and Rouse polymer models predict tracking-FCS curves that are not distinguishable for our particular experimental configuration, possibly because our probe beam's focused waists were too large (so we are less sensitive to the high frequency components of the polymer motion). Other FCS experiments on polymers used waists that were focused nearer to their diffraction limit. In our case it is difficult to control the focused waist because we require that the focal spot be both small and aligned with the axial localization origin of the tracking system, but it is possible that we can improve our focusing with some fine-tuning.

One measurement we may consider is comparing the short-time decay of $g_2(\tau)$ to the power law scalings predicted by the theories and discussed in [24]. It is possible that the long-time decays are less sensitive to the polymer model than these short times, so this is worth examining.

We should also find a reference for calibration of our probe beam via FCS. Most FCS measurements use rhodamine 6G to calibrate the focal volume due to that molecule's very well-known diffusion coefficient, but it has very poor absorption at 415nm, and so is a poor choice for our calibration. We characterized our probe beam by scanning it over the surface of a glass-immobilized fluorescent nanosphere, but it is not clear how that measurement relates to the beam geometry deep inside the liquid sample.

Finally, there are some concerns regarding the label density variation that we observed among our molecules. The exact source of this variation is unknown, as we always followed Invitrogen's labeling protocol (with the exception of a reduced incubation time) when labeling our molecules. Reducing the incubation time could potentially have produced molecules with a lower average label density, but probably not molecules with such variation.

The one conclusion that we can safely draw from our fluorescence data is that we *do* see intramolecular motion of the polymer chain, on time-scales extending out to about 1s. Our measurements are the first to (almost) completely decouple the center of mass and internal diffusive motions of a polymer in three dimensions, which is suggestive of future applications.

Chapter 8

Concluding remarks

With the development of the apparatus described in this thesis, particle tracking microscopy has moved one step closer to viability as a technique for the study of biological systems. The quantum dot tracking demonstrated in Chapter 6 indicates that particles the size of (admittedly, large) proteins are already within reach of our apparatus. Our demonstrations both of heterogeneity in quantum dot fluorescence and of statistical correlations on time-scales longer than the diffusion time are both of a category of measurements that are only possible with tracking techniques. It will be exciting to see what role such techniques will play in biological applications in the future.

The work on DNA polymer dynamics presented in Chapter 7 serves as a reminder that as experimental complexity increases and the boundaries of what we are able to measure expand, it is critically important to be as rigorous in the analysis of measurements as in making the measurements themselves. Early work in this area suffered from both theoretical and experimental deficiencies. While improvements were subsequently made on either the theoretical or the experimental side, our work represents the first merger of rigor on both sides. Our experimental precision allows us to conclude that non-tracking measurements determined biased values for the translational diffusion coefficient of λ-phage DNA and may therefore be difficult to interpret, and that two-dimensional tracking measurements contained more systematic bias than previously thought; our theoretical precision (following that of [24]) allows us to conclude that other tracking measurements on this molecule were analyzed incorrectly; and our measurements suggest that the solvent effects we set out to quantify can be difficult to resolve because, at least in the correlation functions and over the time-scales that we analyzed, no clear signatures of solvent interaction exist or are even predicted in theory. This emphasis of the sensitivity of conclusions to the methods and analysis used to make them is perhaps the most important contribution of this thesis.

Appendix A

Signal analysis via recursive Bayes estimation

A.1 Introduction

Single-molecule confocal fluorescence microscopy has become an important tool for studying the dynamics of diffusing molecules in solution. In this technique, a stream of photons is observed as fluorescent molecules diffuse through a region of a sample that is illuminated by a laser. Current apparatus can measure photon arrival times to sub-nanosecond precision, and observe fluorescence from single molecules in low-concentration samples [105]. This technique has been used to investigate a wide range of chemical and biological problems, both *in vitro* and *in vivo*; however, current methods for the analysis of the fluorescent photon stream work indirectly and are limited in scope. Here we present a rigorous quantitative approach to single-molecule identification, in which we derive a Bayesian estimator and relevant analytical models for the diffusion and fluorescence. The benefit to using analytical models is that they can provide a great deal of insight in complicated experiments. This is not essential for an identification scheme; however, it is our goal here to demonstrate not only an identification scheme but also a general approach applicable to a wide range of problems in single-molecule microscopy. In a more complicated experiment, such as tracking the movement of a molecule or studying dynamics of interactions at the single-molecule scale, only a physical model will suffice to provide rigorous quantitative insight.

Several methods exist for the analysis of the observed photon stream; most generate a particular histogram or autocorrelation function from the stream, and extract information by fitting these to theoretical or simulated functions for the species under examination or from data obtained in reference experiments. These methods were initially developed for estimation of bulk statistical parameters of the sample, such as diffusion coefficient or concentration, though applications in single-molecule experiments have been developed as the technical barriers to detecting single molecules have been overcome [2]. Fluorescence correlation spectroscopy (FCS) analyzes fluctuations in an autocorrelation of the fluorescence intensity signal to extract information about the diffusion coefficient and concentration of the sample molecules, but cannot distinguish between molecules differing in *brightness*, which we define here as the number of bound dyes of a single color per molecule [1, 26, 46]. Alternative formulations of FCS have employed multiple spectral channels and dye colors to enable experiments with multiply-labeled fluorophores [106, 107]. Photon Counting Histogram (PCH) and Fluorescence-Intensity Distribution Analysis (FIDA) measure photon counts in time intervals of fixed duration and fit these measurements to theoretical distributions to differentiate between species with different brightness, but can not differentiate between molecules based on diffusion coefficient [108–110]. Fluorescence Intensity Multiple Distributions Analysis (FIMDA) uses a similar photon-counting histogram but with time intervals of varying duration, and can simultaneously extract information about both diffusion and brightness [111]. The most recent autocorrelation method, Photon Arrival-time Interval Distribution (PAID) combines the autocorrelation approach of FCS with a photon counting histogram to extract information about diffusion and brightness simultaneously, and can be applied to multiple spectral channels [112]. Bayesian estimation has been used previously in identification of multiply-colored fluorophores, however the estimator was simplified considerably in favor of probability distributions obtained by running reference experiments [113]. Additional work in time-correlated single photon counting (TCSPC) has used pulsed excitation lasers and measured the timing of photon detections relative to these pulses to identify molecules based on their diffusion coefficient and has been shown to be more accurate than FCS in single-molecule identification [114, 115]. Here, though, we only consider experiments with a stationary excitation profile.

While the estimation accuracy of the data-fitting methods for bulk sample parameters can be quite good, these methods suffer from several disadvantages when applied to single molecules. Techniques requiring simpler calculations, such as FCS and FIDA, can be applied in real time but are limited in the number of parameters that they can use to distinguish between species [26, 109]. More complex techniques such as PAID are very specific, but the nonlinear fitting algorithms required for data analysis become very computationally intensive and make real-time analysis intractable [112]. In addition, it is difficult to accurately evaluate measurements made from single molecules using a statistical fit because of the infinite number of very different paths that can be taken by a molecule through the laser focus, and the consequently infinite number of possible fluorescence observations. A fitting algorithm will work better for some paths than for others, depending on the similarity between the path taken and the mean value of all possible paths.

Our approach, on the other hand, can be used to distinguish between species differing in virtually any parameter in our model, and is computationally simple in that no fits are required — a probability distribution is directly output at each iteration of the estimator, making real-time signal analysis feasible. It is derived for use specifically in single-molecule experiments, as it includes a spatial component to incorporate variations in the path taken by the molecule through the laser focus. It is not limited to measuring parameters of single molecules, however. As with current methods, accumulated measurements from multiple single-molecule observations can be used to estimate parameters on a larger scale, such as reaction rates or relative concentrations.

The problem that we focus on is the identification of a fluorescent molecule based on the data obtained from a single pass of the molecule through the laser focus, when the set of possible species identities is known. Note that this can be considered a simplification of the parameter estimation problem; to estimate a parameter such as the diffusion coefficient, we simply perform such an identification where the set of possible identities is a set with widely varying values for that parameter. We deal with fluorescence detected from a single spectral channel, although similar approaches are possible with multiple-channel experiments for multiply-labeled species. We derive a Bayesian estimator on the photon stream using distributions for the spatial dependence of the fluorescence rate and the time dependence of the diffusion. We derive analytical expressions for the relevant models for diffusion and fluorescence, rather than relying on Monte Carlo simulations to generate the models. Our estimator is recursively updated after each detected photon, though it may be updated at any desired time interval in the absence of detection, thus maintaining a distribution over the potential set of species inside the laser focus (including the possibility that the focus is empty) at all times.

We present the results of the application of the estimator to simulated two-dimensional diffusion experiments in which molecules are distinguished based on both diffusion coefficient and brightness. Two-dimensional experiments are chosen only for computational simplicity; the filter is derived for diffusion in an arbitrary number of dimensions. Diffusion coefficient and brightness are chosen as the standard parameters used to differentiate between molecules, however our derivations apply to experiments that distinguish between any parameter in our model.

A.2 Derivation of probability distributions

In this section, we will derive the basic filtering equations for identifying the type of a diffusing molecule based on recorded fluorescence photon arrival times. The essential component of this filter is a probability distribution over space and over possible identities of the molecule under observation conditioned on a sequence of fluorescence photon arrival time measurements. We wish to update this distribution in real time as photons are detected, and we assume that at any time at most a single molecule is in the laser focus (this can be ensured by a low sample concentration). By maintaining this distribution, we may specify the most probable identity of the diffusing molecule at any time, given the observations made up to that time.

A.2.1 Recursive Bayesian estimator

Let $S = \{s_1, \ldots, s_n\}$ denote the set of species present in the sample. Let t_k be the time at which the k^{th} photon is detected, and $\xi_n = \{t_1, \ldots, t_n\}$ denote the set of arrival times in a particular experiment up to time t_n. We wish to find an expression for the probability $P(s_j|\xi_k)$, the probability that at time t_k the signal we are observing is due to a molecule of type s_j. Since the fluorescence rate is dependent on the (time-correlated) molecular position $\vec{r}_k \equiv \vec{r}(t_k)$, we must start with a distribution over species and spatial coordinates $p(\vec{r}_k; s_j|\xi_k)$. Using Bayes' rule and the definition of conditional probability, we find the following expression for the probability that a molecule is of type s_j and is at position $\vec{r}_k$ given the observed photon stream ξ_k:

$$\begin{aligned} p(s_j;\vec{r}_k|\xi_k) &= \frac{p(\xi_k|\vec{r}_k;s_j)p(\vec{r}_k;s_j)}{\sum_{j'}\int d\vec{r}\,'_k p(\xi_k|\vec{r}\,'_k;s_{j'})p(\vec{r}\,'_k;s_{j'})} \\ &= \frac{p(t_k|\vec{r}_k;\xi_{k-1};s_j)p(\xi_{k-1}|\vec{r}_k;s_j)p(\vec{r}_k;s_j)}{\sum_{j'}\int d\vec{r}\,'_k p(t_k|\vec{r}\,'_k;\xi_{k-1};s_{j'})p(\xi_{k-1}|\vec{r}\,'_k;s_{j'})p(\vec{r}\,'_k;s_{j'})} \\ &= \frac{p(t_k|\vec{r}_k;\xi_{k-1};s_j)p(\vec{r}_k;s_j|\xi_{k-1})}{\sum_{j'}\int d\vec{r}\,'_k p(t_k|\vec{r}\,'_k;\xi_{k-1};s_{j'})p(\vec{r}\,'_k;s_{j'}|\xi_{k-1})}. \end{aligned} \tag{A.1}$$

In our notation the event t_k together with the measurement record ξ_{k-1} is symbolically identical to the measurement record ξ_k. For diffusion in d dimensions, the integrals in Eq. A.1 are over all of $\mathbb{R}^d$.

We now expand $p(\vec{r}_k; s_j|\xi_{k-1})$ over possible values for $\vec{r}_{k-1}$ to find

$$\begin{aligned} p(\vec{r}_k;s_j|\xi_{k-1}) &= \int d\vec{r}\,'_{k-1} p(\vec{r}_k;\vec{r}\,'_{k-1};s_j|\xi_{k-1}) \\ &= \int d\vec{r}\,'_{k-1} p(\vec{r}_k|\vec{r}\,'_{k-1};\xi_{k-1};s_j)p(\vec{r}\,'_{k-1};s_j|\xi_{k-1}), \end{aligned} \tag{A.2}$$

which contains the recursive term $p(\vec{r}\,'_{k-1}; s_j|\xi_{k-1})$. We substitute into Eq. A.1 to find a formula for the time evolution of the probability distribution over molecular species s_j and position $\vec{r}_k$ conditioned on a set of measured fluorescence photon arrival times ξ_k:

$$p(\vec{r}_k;s_j|\xi_k) = \frac{p(t_k|\vec{r}_k;\xi_{k-1};s_j)\int d\vec{r}\,'_{k-1} p(\vec{r}_k|\vec{r}\,'_{k-1};\xi_{k-1};s_j)p(\vec{r}\,'_{k-1};s_j|\xi_{k-1})}{\sum_{j'}\iint d\vec{r}\,'_k d\vec{r}\,'_{k-1} p(t_k|\vec{r}\,'_k;\xi_{k-1};s_{j'})p(\vec{r}\,'_k|\vec{r}\,'_{k-1};\xi_{k-1};s_{j'})p(\vec{r}\,'_{k-1};s_{j'}|\xi_{k-1})}. \tag{A.3}$$

We were required to carry along a distribution over spatial coordinates in order to develop a recursion formula. In the end, however, we are interested in species identification through the probability $P(s_j|\xi_k)$, which we may now write in terms of Eq. A.3:

$$P(s_j|\xi_k) = \int d\vec{r}_k\, p(\vec{r}_k; s_j|\xi_k). \tag{A.4}$$

We define our species identification estimator $\hat{s}$ as the value of $s_j \in S$ which maximizes Eq. A.4. The estimator $\hat{s}$ defined by Eqs. A.3 and A.4 is exact, in the sense that we have not made any assumptions about the correlation between molecular diffusion statistics and photon detection statistics. In order to update an estimation based on the k^{th} photon detection at time t_k, we only require knowledge of two distributions, $p(t_k|\vec{r}_k; \xi_{k-1}; s_j)$ and $p(\vec{r}_k|\vec{r}_{k-1}; \xi_{k-1}; s_j)$. The former distribution represents the fluorescence emission statistics of a molecule of type s_j at a particular position in the laser focus, while the latter distribution represents the diffusion statistics of a molecule of type s_j. Next, we will describe explicit functional forms for these distributions which are relevant to confocal microscopy experiments.

A.2.2 Effective diffusion statistics

In a typical experimental situation, the motion of a molecule between points $\vec{r}_k$ and $\vec{r}_{k-1}$ is not correlated with any photon detection events. This point is somewhat subtle, and relies on our explicit inclusion of both fluorescence and diffusion statistics. If we detect many photons in a small time interval, it is very likely that a fluorescent molecule is at a position $\vec{r}_k$ of high laser intensity, so that the position of a molecule $\vec{r}_k$ *is* correlated with the measurement record ξ_k. However, the future probability that the molecule will move from position $\vec{r}_{k-1}$ to position $\vec{r}_k$ (without emitting a photon) in time $\Delta t_k \equiv t_k - t_{k-1}$ is independent of the prior photon detections ξ_{k-1}. Symbolically, we have

$$p(\vec{r}_k|\vec{r}_{k-1}; \xi_{k-1}; s_j) = p(\vec{r}_k|\vec{r}_{k-1}; \Delta t_k; s_j), \tag{A.5}$$

where the right-hand side of Eq. A.5 is the usual Green's function solution to a Fokker-Planck equation for the diffusion statistics of molecules of type s_j.

In most cases of experimental interest, we may average over three-dimensional distributions to obtain a set of estimator equations which is effectively two-dimensional. Furthermore, if we consider a cylindrically symmetric laser excitation profile, we may reduce all of the (vector) coordinates $\vec{r}_k$ to (scalar) cylindrical radial coordinates r_k, where as before $r_k \equiv r(t_k)$. For isotropic, force-free Brownian motion projected into $d = 2$ dimensions, we can solve Eq. A.5

analytically to find

$$p(r_k|r_{k-1};\Delta t_k;s_j) = \frac{1}{4\pi D_j\Delta t_k} I_0\left(\frac{r_k r_{k-1}}{2D_j\Delta t_k}\right)\exp\left(-\frac{r_k^2+r_{k-1}^2}{4D_j\Delta t_k}\right), \tag{A.6}$$

where D_j is the diffusion coefficient for molecules of type s_j and I_0 is the zeroth-order modified Bessel function of the first-kind.

A.2.3 Fluorescence photon detection statistics

Equation A.6 contains all of the information we need to know about diffusion in order to implement the filter, Eq. A.3. Next we will develop an explicit expression for the fluorescence photon detection statistics $p(t_k|r_k;\xi_{k-1};s_j)$ (we have dropped the vector notation on r_k as before). For simplicity, we use a simple two-level, saturating emitter model of fluorescence, but the implementation of the filter is essentially the same for any rate-equation model or model incorporating the internal dynamics of the fluorophore [17]. In this model, the instantaneous rate $\gamma_j(r_k)$ of photon emission by a molecule of type s_j at position r_k and labeled with m_j (identical) dye molecules is determined by the spatially-dependent laser excitation rate $\Gamma_L(r_k)$, the relaxation rate of the dye molecule Γ_j and the background noise count rate Γ_B:

$$\gamma_j(r_k) = m_j\frac{\Gamma_L(r_k)\Gamma_j}{\Gamma_L(r_k)+\Gamma_j} + \Gamma_B. \tag{A.7}$$

$\Gamma_L(r)$ is proportional to the laser beam intensity, which we take to be Gaussian with beam waist w in the transverse dimensions, and we neglect the variation of the excitation intensity in the axial direction:

$$\Gamma_L(r) = \Gamma_L(0)\exp\left(-\frac{r^2}{2w^2}\right). \tag{A.8}$$

It should be noted here that finite efficiency photon detection does not affect the forms of the filter or the fluorescence model. We can always scale the rates in Eq. A.7 by the detection efficiency η which has no effect on the spatial dependence of fluorescence statistics (except to decrease the overall rate of photon detection). Furthermore, for small η, we may assume that a fluorophore emits many photons ($\sim 1/\eta$) between any two photon detections. This ensures that the fluorophore is in its radiative steady-state and we may safely neglect any quantum statistics associated with single-molecule photon detection. Since the form of Eq. A.3 is largely independent of η, we often neglect the distinction between photon *emission* and *detection* rates, whose ratio is η by definition.

The probability $2\epsilon\, p(t_k|r_k;\xi_{k-1};s_j)$ is the likelihood of a photon detection in the infinitesimal interval $t_k \pm \epsilon$ with no other photons detected at times $t \in (t_{k-1}, t_k)$. This probability can be expressed as

$$2\epsilon\, p(t_k|r_k;\xi_{k-1};s_j) = 4\pi\,\epsilon\gamma_j(r_k)\exp\left(-2\epsilon\gamma_j(r_k)\right) \\ \times\int dr_{k-1}\,r_{k-1}\,p(n_k=0|r_k;r_{k-1};\xi_{k-1};s_j)p(r_{k-1}|\xi_{k-1};s_j), \quad \text{(A.9)}$$

where n_k is the number of photons emitted in time interval Δt_k. Note that the precise value of the factor ϵ is unimportant because $\exp\left(-2\epsilon\gamma(r_k)\right) \approx 1$ since ϵ is very small by definition. The integral represents an average over all possible starting radial positions r_{k-1} from which the molecule moves to radial position r_k in the time interval Δt_k. In general, n_k depends on the path taken by the molecule over the time interval, which makes calculating an exact analytical expression for Eq. A.9 difficult, although a path integral representation is possible (see Appendix A.5). The difficulty in calculating this function arises from the variation of the photon emission rate over the possible paths that a molecule can take from r_{k-1} to r_k. We expect the fluorescence count to obey Poisson statistics, so

$$p(n_k=0|r_k;r_{k-1};\xi_{k-1};s_j) \approx \exp\left(-\gamma_j(r_k)\Delta t_k\right) \quad \text{(A.10)}$$

as long as a molecule does not move between photon detections to a position of very different excitation intensity. We expect the approximation in Eq. A.10 to hold for $w^2\gamma_j(0) \gg D_j$, so that a molecules does not move too far between photon detections. This is precisely the experimental regime of interest, since this condition is violated when few photons are detected in a single-molecule transit through the laser focus, and we would not expect to gain much information in this case anyway. In Appendix A.5 we calculate the first-order correction to Eq. A.10. For typical parameters in our simulations, the numerical value of this correction factor is less than 2% for regions within the beam waist of the laser.

A.2.4 Practical Considerations

Several details must be considered before the filter can actually be implemented. First, while the integrals in Eq. A.3 are over all of $\mathbf{R}^d$, we must truncate them numerically. We set the integration limit $R_{max} > w$, the laser beam waist, so that $\Gamma_L \approx 0$ in the regions excluded from the integrals; in practice, we find that $R_{max} = 4w$ is sufficient.

Next, to account for the loss of probability at the boundary R_{max} due to the diffusion term, we add a distribution representing the probability of molecules from the outer regions diffusing into the region bounded by R_{max} to the spatial distribution in Eq. A.3 at each iteration of the estimator. This takes into account the concentrations of the different species in the sample, and is given by

$$p_{leak}(r_1|\xi_k;s_j) = 2\pi C_j \int_{R_{max}}^{\infty} dr_0\, r_0\, p(r_1|r_0;\xi_k;s_j) \tag{A.11}$$

which is a convolution of Eq. A.6 with the distribution of molecules in the outer region, which we assume to be uniform with concentration C_j. This leak term allows us to run the estimator in the absence of a molecule in the laser focus. To account for the possibility of an empty laser focus, we include a dark species s_0 with brightness $m_0 = 0$ in all iterations of the estimator. Note that we have not yet defined the initial distribution $p(s_j;\vec{r}_0)$; in order to obtain this distribution we simply run the estimator on background noise, in the absence of fluorescent species, until a steady-state distribution is reached.

Finally, we must decide how to determine when a molecule has either entered or exited the laser focus. For this we use the spatial distributions generated by the estimator. We decide on a particular threshold radius R_{th} that bounds what we consider to be the laser focus, and calculate the probability that the focus is empty,

$$P(empty|\xi_k) = P(s_0) + 2\pi \sum_{j\neq 0} \int_{R_{th}}^{R_{max}} dr_k\, r_k\, p(r_k;s_j|\xi_k). \tag{A.12}$$

We decide on a minimum value P_{low} for this probability and consider a molecule to have entered the laser focus when Eq. A.12 drops below this value. Additionally we choose a maximum value P_{high} and consider the molecule to have exited the focus when Eq. A.12 increases above P_{high}. At this point, we integrate the spatial distributions to obtain an estimate of the identity of the observed molecule. We leave detailed discussion of the motivation for this scheme to a future publication.

With these details specified, our formulation of the estimator is complete. Next, we discuss the expected strengths and limitations of the estimator under particular experimental conditions.

A.2.5 Experimental regimes

While the derivation of Eq. A.3 is applicable to all cases of molecular diffusion and fluorescence, we expect its performance to be affected by the parameters of the experiment. For instance, we stated that Eq. A.10 is a poor approximation when the diffusion rate is very fast relative to the fluorescence rate $\gamma_j(\vec{r})$. We may attempt to evaluate Eq. A.9 more accurately, but the performance of our estimator will still be quite limited due to its dependence on a spatial distribution that changes more rapidly than it is updated. We expect the filter to be most effective when $\gamma_j(\vec{r})$ is high and diffusion rate is low, so that the value we calculate for $p(\vec{r}_{k-1};s_j|\xi_{k-1})$ remains a good estimate for $p(\vec{r}|\xi_{k-1};s_j)$ as the molecule moves from $\vec{r}_{k-1}$ to $\vec{r}_k$.

It may be the case that the diffusion rate is very small relative to γ_j — such is the case for diffusion of molecules on a membrane, for example. In this case we expect to detect large numbers of photons at a fairly constant rate, so that little spatial information is gained from the time spacing between photons. Hence we expect that we can ignore this spacing and the exact arrival-time data, and instead count the numbers of photons detected in individually-spaced time bins of arbitrary size τ_{bin}. Then we replace $\xi_k = \{t_1, \ldots, t_k\}$ with $\xi_k = \{n_1, \ldots, n_k\}$ in deriving Eq. A.3, where we define n_k as the number of photons counted in the time bin (t_{k-1}, t_k). We define $\vec{r}_k = \vec{r}(t_k)$ as before. Here the diffusion probabilities $p(\vec{r}_k|\vec{r}_{k-1}; \xi_{k-1}; s_j)$ are calculated as in Eq. A.5, and the fluorescence probabilities are calculated using

$$p(n_k|r_k; \xi_{k-1}; s_j) \approx \exp\left[-\tilde{\gamma}_j(r_k)\tau_{bin}\right] \frac{\left[\tilde{\gamma}_j(r_k)\tau_{bin}\right]^{n_k}}{n_k!} \tag{A.13}$$

where the approximation, as in Eq. A.10, comes from the molecule's path-dependent fluorescence rate.

This formulation of the estimator requires significantly fewer computations, because the diffusion and fluorescence probabilities can all be calculated prior to the experiment, making real-time estimation easier to achieve. We expect the performance to be determined in part by the bin time τ_{bin} that we choose — a τ_{bin} that is too large ignores significant diffusion dynamics, but more computations are required as the bin time is made smaller — so the choice of bin time requires balancing these tradeoffs.

A.2.6 Generalizing the approach

We have stated that the most important point that we are presenting here is a general approach to computational single-molecule studies. It is important, then, to consider how our approach generalizes to other experiments. The recursive estimator given by Eq. A.3 is a valid form for any experiment in which no more than a single molecule is likely to be in the focus of the laser; modifications or extensions of the models for diffusion and fluorescence do not affect the form of the estimator. Obvious extensions to the fluorescence model may incorporate multiple spectral channels to facilitate further identification or experiments using fluorescence transfer [17]. Additionally, more fluorescence details may be incorporated if necessary, such as dye photobleaching, blinking dyes, forbidden transitions, and spatially-dependent photon collection efficiency. Extensions to the diffusion model can be made to incorporate diffusion restricted to a particular spatial domain, diffusion with net flow, and as we have stated, free diffusion in higher or lower dimensions. Clearly this makes feasible the use of this recursive estimator in many common chemical and biological experiments.

As stated earlier, parameter estimation can be treated as an extension of identification as we have defined it here. Any parameter that can be incorporated into our model for identification can also be approximated by our technique if we simply perform identification over a set in which that parameter is varied — this is precisely a maximum likelihood estimator [see 116, for example]. It is also possible to use a similar approach to derive estimators for parameters that are relevant on the bulk sample scale, not the single-molecule scale. For example, the details of single-molecule observations are not affected by sample concentration, though the frequency of such observations is. It is not very difficult to write a concentration estimator using a model for this frequency. Coupled with the identification estimator, this could be used to estimate the concentrations of the different components of a sample. Thus all of the functionality of FCS and other current techniques can be achieved by rigorous estimation and extended to include any exotic photophysical or diffusion dynamics for which a suitable model exists.

Our ultimate goal is the development of techniques that facilitate experiments in which exotic dynamics will limit the efficacy of current methods. Current techniques are limited to experiments for which sufficient autocorrelation functions or histograms can be generated for fitting to the experimental data. We envision experiments in which fluctuations in the quantities being measured affect the fluorescence stream in such a way that realizing accurate results by fitting observations to a small number of sample functions will be impossible. For example, to study time-dependent mechanical oscillations or rotations in single molecules — such oscillations are common to proteins yet current techniques for their study are quite limited [117] — an FCS experiment would require an enormous number of model autocorrelations to explore a significant portion of the infinite-dimensional space that such oscillations lie in. In contrast, with an appropriate model our recursive update approach would be capable of making a statistically best estimate of oscillatory state of such a system continuously in time for as long as the system can be observed. We can also imagine the study of random events that have a dramatic effect on diffusion dynamics, such as active inter-compartmental transport of free molecules in a living cell, in which a there is not only a random short-duration change in the diffusion coefficient during transport, but also a change in the topology of the space over which the molecule is confined. Again, for such a process it will be very difficult to calculate or simulate a sufficient set of functions to fit the observations to, while a recursive update formula with a sufficient model can be readily written down. We believe our approach is the best path to take for the study of such complex stochastic single-molecule dynamics.

Now we will present the results of simulations to demonstrate and characterize the estimator in the single-molecule identification experiment.

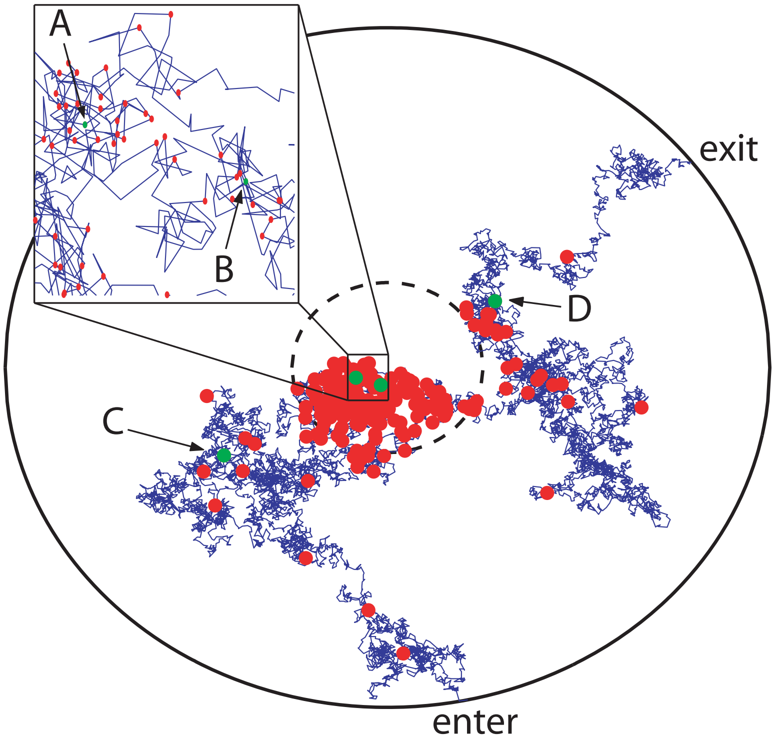


Figure A.1: Data generated by two-dimensional simulation. The plot shows a trajectory of a molecule with $D_j = 0.25\mu\text{m}^2/\text{ms}$ and $m_j = 1$ through the circular region centered at the laser focus with radius $R_{max} = 2\mu\text{m}$. The inner, dashed circle represents the laser beam waist. Entry and exit points are indicated. Displayed molecule positions are updated in 0.5μs increments. Circles indicate the molecule's position at each photon arrival time. Labeled points correspond to the times at which the distributions in Fig. A.3 are shown.

A.3 Simulations

In this section, we will simulate diffusion experiments in two dimensions in MATLAB to illustrate the use of the recursive filtering technique for the analysis of photon streams. The simulations specify a set S of species with varied diffusion coefficients D_j and brightness m_j, and all other model parameters held constant. Discrete, fine-grained ($\Delta t = 0.5\mu$s) Brownian motion trajectories are generated for a particular species, and photon streams are generated from a Poisson distribution with rate given by Eq. A.7. We set the beam waist $w = 0.5\mu$m and rates $\Gamma_L(0) = 637\text{ms}^{-1}$ and $\Gamma_j = 500\text{ms}^{-1}$ for all j. We first set $\Gamma_B = 1$kHz to evaluate the estimator in the presence of minimal background noise. The arrival times of photons are extracted from the stream and sent to the estimator for evaluation, and identifications are made from the resulting distributions using $R_{max} = 2\mu$m, $R_{th} = w = 0.5\mu$m, $P_{low} = 0.25$ and $P_{high} = 0.75$.

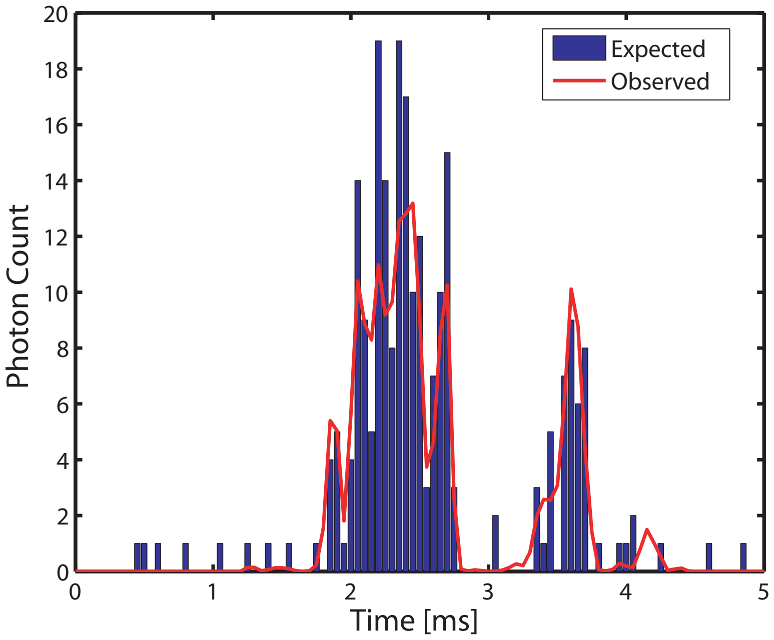


Figure A.2: Simulated photon counts detected in time intervals of 50 μs as the molecule traversed the path in Fig. A.1. Also shown is the expected photon count over these intervals based on the actual laser intensity at the molecule's position, given by Eq. A.7. Note the difference between expected and detected fluorescence rates — this contributes to the difficulty of extracting accurate spatial information from the photon stream.

A.3.1 Identification based on diffusion coefficient

An experiment that we may perform using our identification algorithm is one in which the dye-labeled sample molecules may polymerize with unlabeled molecules, and we try to distinguish between monomers and dimers. For example, such an experiment could be used to distinguish between single-stranded and double-stranded DNA. Figs. A.1 and A.2 show a sample trajectory and photon stream generated for an experiment with $S = \{(1, 0.25), (1, 0.5)\}$, where we denote $s_j = (m_j, D_j)$ with D_j in units of $\mu\text{m}^2/\text{ms}$; these values are consistent with small nucleic acid polymers [91, 112]. The trajectory shown is for a molecule of type s_1. This photon stream was run through the estimator and some of the resulting spatial distributions are shown in Fig. A.3. We see in these distributions the effect that random fluctuations can play in the estimate; while the peaks of the spatial distributions for the molecule at positions A, B, and D are nearly exactly at the molecule's actual radial position, the estimator is off by nearly one beam waist when the molecule is at position C. Statistically, such momentary inaccuracies are expected; by its recursive nature, however, the estimator corrects such errors quickly as additional photons are detected.

To evaluate the performance of the recursive estimator in this type of experiment over a range of possible S, we vary the ratio of diffusion coefficients between species: we set S =

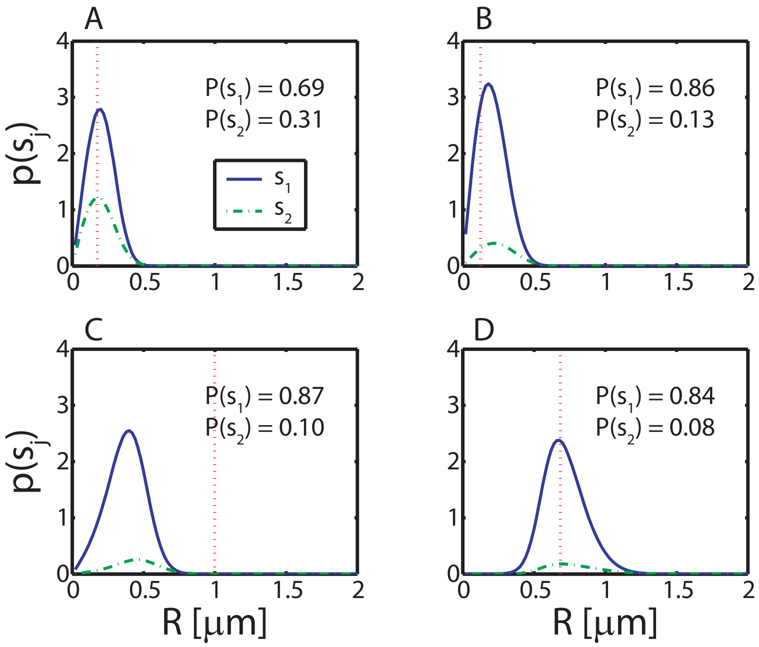


Figure A.3: Sample spatial distributions resulting from the stream of 253 photons presented in Fig. A.2, using $S = \{(1, 0.25), (1, 0.5)\}$. The distribution after the 62^{nd} photon detected (point A in the trajectory) is in the top left. The top right, bottom left, and bottom right distributions are after the 125^{th}, 188^{th} and 252^{nd} photons, with the molecule at positions B, C, and D, respectively. Integrated probabilities given by Eq. A.4 are reported for each species. Dotted vertical lines indicate the actual radial position of the molecule.

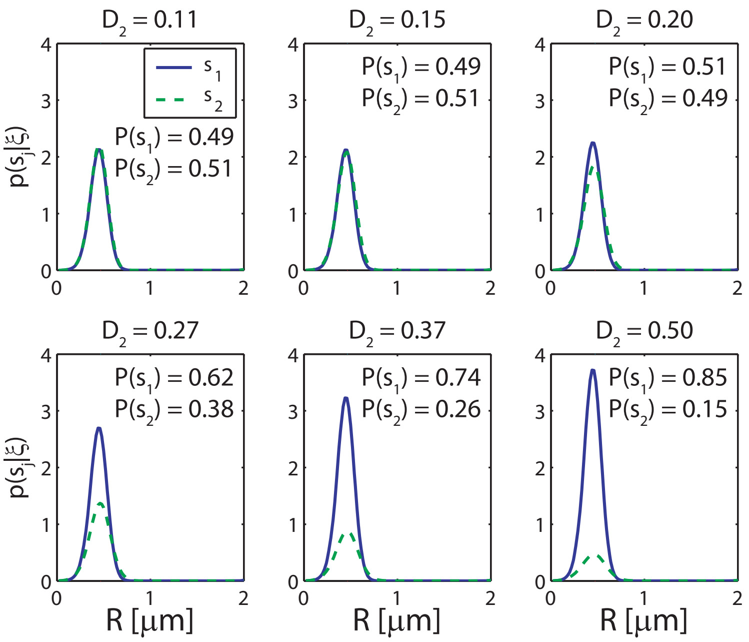


Figure A.4: Performance of recursive estimator in identifying species with varied diffusion coefficient ratios. One trajectory and resulting set of 91 photon arrivals was generated for molecule type s_1 with the set $S = \{(1,\ 0.1),\ (1,\ D_2)\}$, for 6 logarithmically-spaced values of D_2 . Shown are the radial distributions for each species after the final photon arrival time was processed. Integrated probabilities are indicated.

$\{(1, 0.1), (1, D_2)\}$ with $0.1 \leq D_2 \leq 0.5$. Fig. A.4 shows the final probability distributions over space and species type for a single simulated trajectory and set of photon arrival times for a molecule of type s_1, run through the estimator with varied values for D_2. This highlights an important aspect of the estimator: in addition to providing an identification scheme, it provides a quantitative evaluation of the accuracy of the identification in the form of the probability that the estimated species is indeed present. The estimator was not able to correctly identify the molecule for $D_2 \leq 0.15$; however, as D_2 increased to 0.5, $P(s_1|\xi)$ increased to 85%.

We now consider the accuracy of the estimator in identifying molecules relative to theoretical upper and lower performance bounds. A primary difficulty in extracting information about the diffusion dynamics lies in the uncertainty in estimating molecule position from fluorescence rate, since the detected fluorescence rate is not a deterministic function of position (see Fig. A.2). In theory, the best possible identification estimate could be made if the exact position of the molecule could be extracted from the data at each iteration of the estimator. To simulate this we generate trajectories and photon streams in a manner identical to that used for the recursive estimator, but we record the exact radial position of the molecule at every photon arrival time. We use a maximum likelihood estimator [see 116, for example] for species identity from the position and arrival-time data; we take the probability of successful identification by

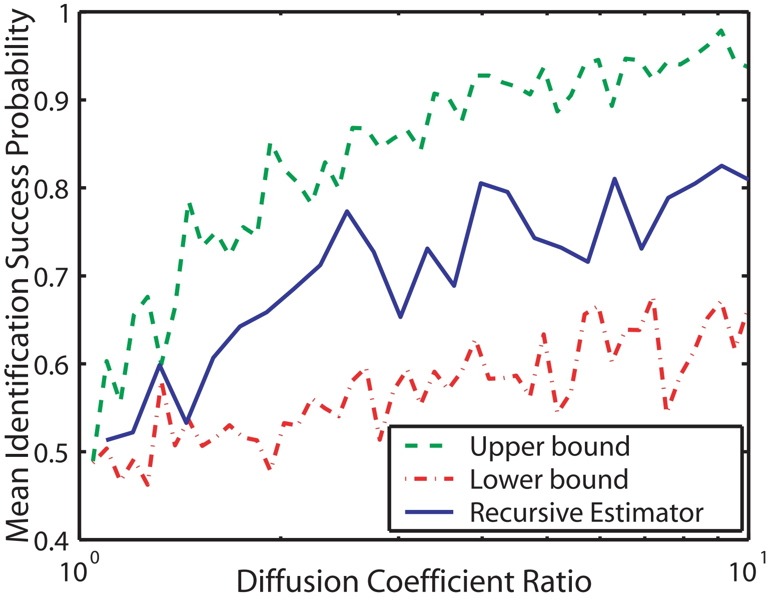
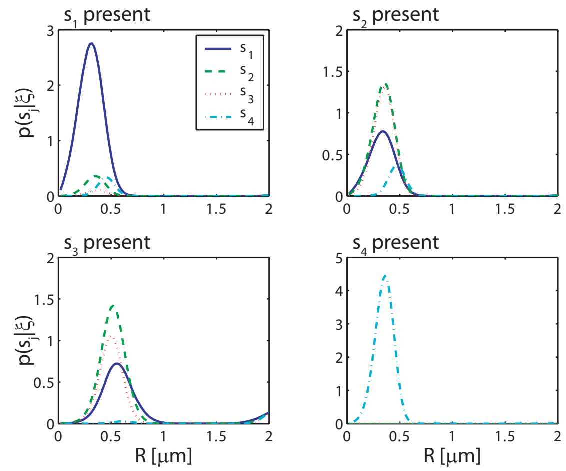


Figure A.5: The performance of the recursive estimator as a function of diffusion coefficient ratio, compared to that of the estimators representing theoretical upper and lower performance bounds. We set $S = \{(1, 0.1), (1, D_2)\}$ and generate trajectories resulting in at least 200 identifications for each species type, for each estimator, for varied D_2. Plotted are the geometric mean success probabilities for each estimator.

this method to be the theoretical upper bound for performance in any simulated experiment. In addition, we decide on a lower bound for the success probability by considering only the duration of the identification trajectory — this should contain the least possible information about the diffusion dynamics. We numerically approximate $p(\tau|D_j)$, the probability distribution of the time τ spent within the laser focus given the diffusion coefficient D_j, and let our identification estimate for each trajectory be the species for which this probability is greatest. We take the performance of such an estimator to be the minimal performance that we should be able to achieve with data from a single trajectory.

Fig. A.5 compares the accuracy of the recursive estimator to that of the theoretical upper and lower bounds. For clarity, success probabilities are not shown for each species, but rather the geometric mean of the two probabilities is shown. The geometric mean was chosen as a simple measure of both the accuracy and the bias of the estimator; it is highest only when the estimator is nearly unbiased and accurate for both species. For diffusion coefficient ratios less than 1.5, the probability of successful identification lies near 50% for all three estimators. For ratios above 1.5, the recursive estimator lies well within the bounds. Note that here we only consider the probability that the estimator correctly reports the species present, not the reported probability of the estimate. We expect to be able to improve the performance by specifying a minimum threshold for this probability and thereby only accepting estimates that



Species Present	P (s_1)	P (s_2)	P (s_3)	P (s_4)
s_1	0.82	0.09	0.03	0.6
s_2	0.24	0.36	0.32	0.08
s_3	0.27	0.42	0.29	0.02
s_4	0.00	0.00	0.00	1.00

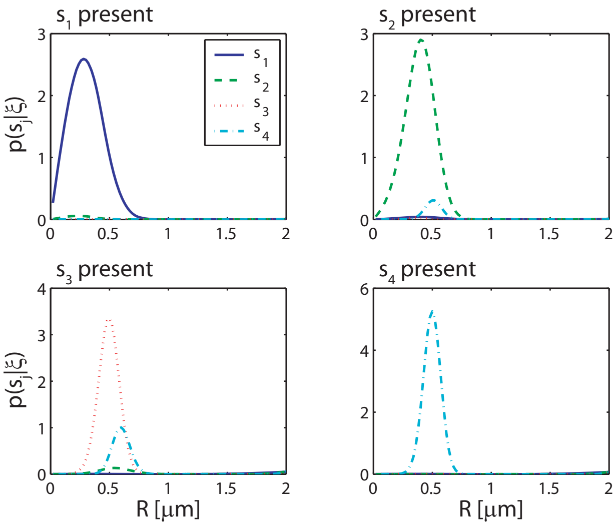
Figure A.6: **left:** The performance of the recursive estimator in the multiple component experiment with $S = \{(1, 0.4), (1, 0.2), (1, 0.15), (2, 0.15)\}$. Trajectories were generated for each species, and the resulting photon arrival-time data was run through the estimator. The distributions shown are the distributions prior to the final exit time of the molecule for each trajectory. **right:** The table shows the integrals of the probability distributions shown in the figure, giving the probability of each species being present according to the data generating each plot.

are made with a high degree of certainty; these data represent a minimal level of performance for our estimator.

Identification of diffusing fluorophores by diffusion coefficient alone is a problem addressed by FCS. Using this method, a nonlinear fit to simulated data or a maximum likelihood estimator is used to identify a species given its autocorrelation function [114]. Our technique has several advantages over this method, in that the estimator provides a rigorous probability distribution over species present in the sample. Thus, in addition to a most probable identity, it provides a measure of how certain the estimate can be considered to be. In addition, such an estimate can be made at any time during the fluorophore's path through the beam focus, allowing an experiment to interact with sample molecules before they leave the focus, as is necessary for a tracking or sorting experiment. FCS methods, in contrast, rely on an autocorrelation that may not provide a good identification estimate with incomplete data. Of course, a major advantage over FCS is that we can consider additional molecular parameters in our identifications. Next we consider an experiment in which the species present may differ in brightness m_j.

A.3.2 Multiple species identification

Here we propose a hypothetical experiment in which four dye-labeled species exist in solution: one small molecule, one large molecule, one small-large molecule complex, and one small-large complex where one of the dyes has photobleached. Such experiment could be useful if, for instance, we wanted to measure the kinetics of complex formation. We can also imagine an



Species Present	$P(s_1)$	$P(s_2)$	$P(s_3)$	$P(s_4)$
s_1	0.98	0.02	0.00	0.00
s_2	0.02	0.92	0.01	0.05
s_3	0.02	0.05	0.74	0.20
s_4	0.02	0.00	0.00	0.98

Figure A.7: **left:** The performance of the recursive estimator in a high diffusion coefficient ratio multiple component experiment with $S = \{(1,2),(1,0.4),(1,0.08),(2,0.08)\}$. Trajectories were generated for each species, and the resulting photon arrival-time data was run through the estimator. The distributions shown are the distributions prior to the final exit time of the molecule for each trajectory. **right:** The table shows the integrals of the probability distributions shown in the figure, giving the probability of each species being present according to the data generating each plot.

experiment using real-time feedback to isolate only the complex with two intact dyes from solution. To simulate such experiments, we let $S = \{(1,0.4),(1,0.2),(1,0.15),(2,0.15)\}$ to test low diffusion coefficient ratios and $S = \{(1,2),(1,0.4),(1,0.08),(2,0.08)\}$ to test high ratios. The ratios between molecular diffusion coefficients of complexes and their constituents are largely dependent on the relationships between the geometry of the individual components and that of the complex. Hence both cases have some physical relevance, with the low-ratio case an example where geometry is largely not altered by binding, and the high-ratio case an example in which the individual constituents may be tightly folded, for instance, while the complex takes a large, extended form.

Figs. A.6 and A.7 show distributions resulting from trajectories generated by each species type in the multiple species experiments. In the low diffusion coefficient ratio experiments, the estimator identifies s_1 and s_4 with high probability, but cannot distinguish well between s_2 and s_3 due to their very similar diffusion coefficients. For higher diffusion coefficient ratios, all species are identified with high probability. Note that in both cases, s_4 is identified with probability nearly 1, indicating that brightness is a much stronger criterion for distinguishing between species than diffusion coefficient.

Experiments to distinguish between species based on both diffusion coefficient and brightness are the focus of the FIMDA and PAID methods [111, 112]. Our method has several advantages over these. Again, our estimator reports identification probability, a rigorous measure of

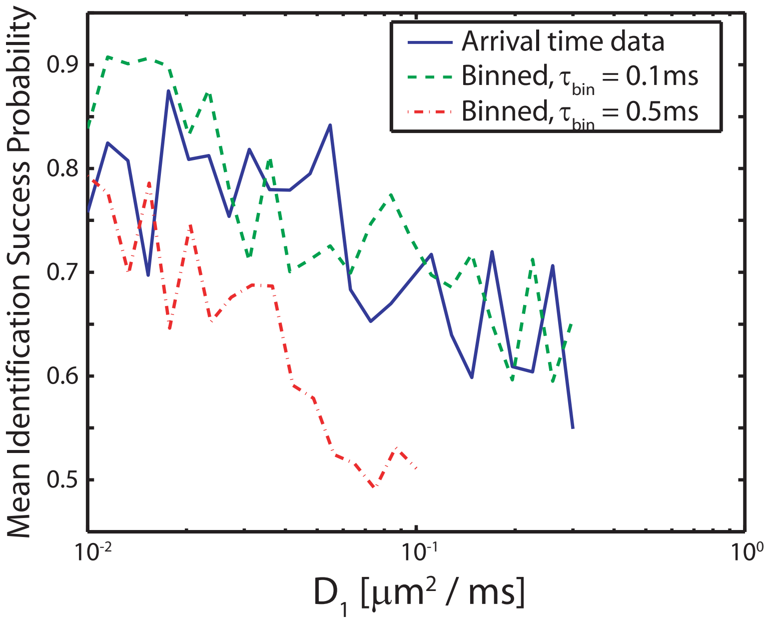


Figure A.8: Comparison of estimator performance using binned photon-count data and arrival-time data. Geometric mean success probabilities are shown for simulations in which we set $S = \{(1, D_1), (1, 3D_1)\}$ and vary D_1. A minimum of 100 identifications were made for each species, for each value of D_1, for both estimators. Binned data were generated with bin times $\tau_{bin} = 0.1$ms and $\tau_{bin} =$0.5ms, as indicated in the legend. For $\tau_{bin} = 0.5$ms, the estimator cannot successfully observe species s_2 for $D_3 > 0.3\mu\text{m}^2/\text{ms}$, so the plot does not continue beyond this point.

the confidence in the estimate. This is useful in any experiment — for kinetics experiments, we can calculate the appropriate uncertainties in whatever values we measure; for sorting, we can be highly certain that we extract the correct molecules. Second, FIMDA and PAID require multidimensional nonlinear fits of their data to simulated or theoretical results, making real-time experiments nearly impossible with modern computer hardware. Third, as in the comparison with FCS, our technique always provides the statistically-best estimate of the species present at all times, providing information for real-time experiments while the fluorophores are within the focal volume. While we are not aware of any studies that examine the performance of FIMDA and PAID in the identification of single diffusing molecules, we expect that our estimator should be at least as accurate as these methods.

A.3.3 Slow-diffusion identification

In experiments where species diffusion coefficients are small relative to the fluorescence rates, we expect that a reduction of the photon stream from a series of arrival times to a series of photon counts in time windows of length τ_{bin} will be both effective and computationally simpler than the arrival-time estimator that we have been using. We test this by running a series of two-species diffusion-based identification simulations, setting $S = \{(1, D_1), (1, 3D_1)\}$ and varying D_1. Fig. A.8 shows the results of these simulations, run using bin times of 0.1ms and 0.5ms. We see that for small diffusion coefficients, with sufficiently small bin time, the binned-data estimator's performance reaches as high as 90% — roughly ten percentage points higher than the performance of the arrival-time estimator. For $D_1 \geq 0.02\mu\text{m}^2/\text{ms}$, both estimators achieve roughly the same performance. When a larger bin time is used, the performance of the binned-data estimator is degraded, falling as low as 50% before the data become too coarse-grained to make identifications, which happens for $D_2 > 0.3\mu\text{m}^2/\text{ms}$.

It is somewhat surprising to see the binned-data estimator outperform the exact arrival-time estimator. We attribute this result to accumulated numerical error as many photon arrival times are processed; finite precision mathematics (MATLAB uses 64-bit data types) limits the accuracy of each iteration of the estimator. The binned-data estimator may update as little as 10% as frequently as the arrival-time estimator, and as a result the difference in accumulated error is great enough that the binned estimator performs better despite the arrival-time estimator's theoretical advantage. The effect is particularly pronounced for molecules with small diffusion coefficients because a very large number of photons is typically detected as such molecules diffuse through the laser focus. This results in a large number of recursive updates, each introducing some numerical error to the distribution.

As shown in Fig. A.8, the performance of the binned-data estimator will be limited by selection of the bin time; to achieve best results, the bin time must be made sufficiently small to capture essential features of the fluorescence stream, yet not too small to cause numerical error to hurt accuracy. Initial speculation about the implementation of the recursive estimator on a Field Programmable Gate Array [118] leads us to believe that the binned-data estimator with reasonable bin time reduces the frequency of computations sufficiently to enable its use in experiments requiring real-time feedback.

Clearly the binned-data estimator, not the arrival-time estimator, is the best tool for the current problem both for computational efficiency and accuracy, at least when numerical precision is limited to 64-bit. However, the arrival-time formulation will be essential for experiments in which fast photophysical dynamics, such as blinking or energy transfer, are incorporated into the fluorescence model. The binned-data estimator will never be able to accurately handle dynamics on a time scale smaller than τ_{bin}, while the arrival-time estimator's time resolution is limited only by photodetector resolution.

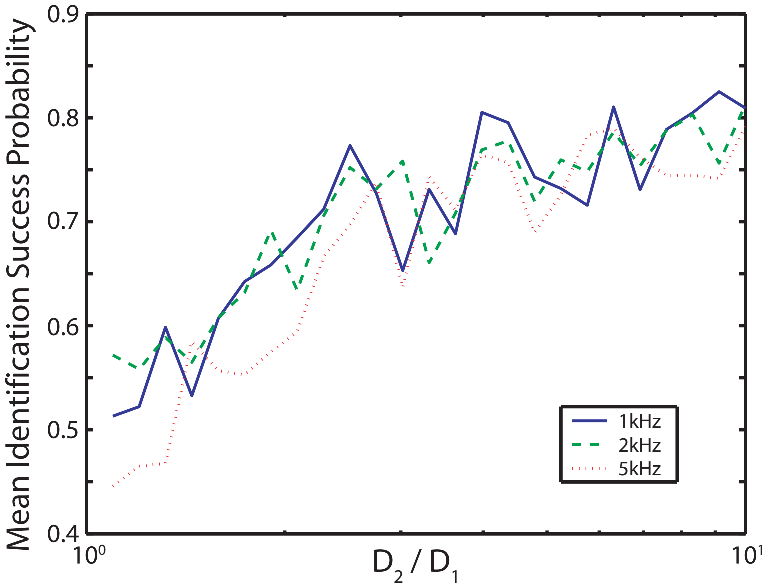


Figure A.9: Performance of the arrival-time estimator on data with varied background noise strengths indicated. The plots were generated using $S = \{(1, 0.1), (1, D_2)\}$, $0.1 < D_2 \leq 1.0$, with a minimum of 200 identifications made for each species. Data shown are the geometric means of the success probabilities for each species.

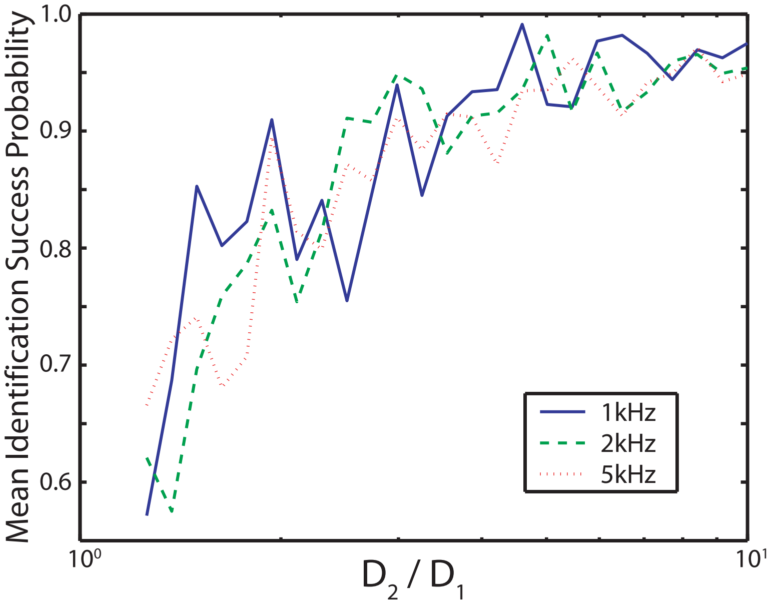


Figure A.10: Performance of the identification estimator on binned data with bin time 0.1ms and varied background noise strengths as indicated. The plots were generated with $S = \{(1, 0.01), (1, D_2)\}$, $0.01 < D_2 \leq 0.1$, with at least 100 identifications made for each species. Shown are the geometric means of the success probabilities for each species.

A.3.4 Background noise and estimator performance

So far we have only presented results showing the performance of the estimator with 1kHz background noise. Here we consider the effect of higher-rate background fluorescence on the estimator. The model that we derived for the fluorescence incorporated a constant-rate background count Γ_B. We expect that Eq. A.7 will provide fluorescence rates that are exact assuming that the noise has a constant rate that can be measured or approximated prior to the experiment.

Figs. A.9 and A.10 show the identification accuracy of the estimator on data containing constant-rate Poisson background noise. The plots are generated by setting $S = \{(1, D_1), (1, D_2)\}$ with D_1 constant at 0.1 for the arrival-time data and 0.01 for binned data, and varying the ratio D_2/D_1. Note that a direct comparison between the probabilities in these figures is irrelevant because the plots are over different diffusion coefficient regimes. We see that there is little noticeable difference in the performance of the estimator on either data type with noise at 1kHz and 2kHz, but for noise strengths of 5kHz there is a performance loss of a few percentage points for the arrival-time data. The difference in noise-rejection performance between the data types is again attributed to numerical issues: higher noise count rates increase the number of iterations of the arrival-time estimator and therefore increase propagated error. Typical background count rates in single-molecule fluorescence experiments are about 1-2kHz, and both estimators are robust to noise at these levels.

A.4 Conclusion

We derived a recursive Bayesian estimator to calculate a probability distribution for the identity of a single diffusing fluorophore given its photon arrival-time stream as it makes a single pass through the focus of a confocal fluorescence microscope. We derived analytical models for the diffusion and fluorescence dynamics in such an experiment, and tested our estimator by running experimental simulations in two dimensions.

We showed that our estimator is capable of identifying single molecules based on differences in their diffusion coefficients and brightness, but we stress that our method provides a means of identification that based on any species-specific parameter. Our simulations indicated that the accuracy of the estimator in identification by diffusion coefficient is reasonable, as it lies within the performance bounds set by a minimal-data estimator and an ideal, theoretical one. We showed that our estimator is significantly more sensitive to brightness than it is to diffusion coefficient by demonstrating that species with twofold differences in brightness can be identified with probability 1, while estimator accuracy on species with twofold differences in diffusion coefficient is not much better than random guessing.

A key feature of our identification scheme is that it provides a probability distribution over the set of possible species identities. This provides a measure of the certainty of an identification, as we demonstrated that relatively similar species are identified with fairly low probability, and relatively dissimilar ones are identified with high probability.

We showed that by counting photon arrivals in time windows of fixed duration we can significantly reduce the number of calculations necessary while maintaining accuracy in diffusion coefficient-based identification for slowly diffusing molecules. In fact, accuracy improved by doing this because of issues with the accumulation of numerical errors over many estimator iterations. We believe that this approach is experimentally tractable, in that it is both accurate and computationally simple enough to facilitate experiments requiring real-time feedback.

It is important to note that our work was done in two dimensions purely for computational simplicity and to illustrate the technique. The estimator is valid in an arbitrary number of dimensions, and the models we derived are easily extended to three dimensions. Assuming a correct model, we believe that the performance of the estimator applied in three dimensions will not be much different than that which we have presented.

The most important point we wish to address is the prospect of using approaches similar to what we have presented to tackle more complicated problems in single-molecule spectroscopy. Rigorous mathematical tools exist for the treatment of stochastic processes such as diffusion and fluorescence, and when applied properly they yield direct quantitative approaches to the same problems that have previously been addressed using more limited and circuitous methods. These quantitative tools are only limited in the extent to which the physical processes underlying the experiments are understood; for any system that we can write an accurate model for, we can develop an estimator to measure something about that system. We hope we have made clear the motivation for our belief that coupling a rigorous quantitative approach to data analysis with clever experiments will allow for novel studies of complex stochastic dynamics on the single-molecule scale.

A.5 Perturbative calculation of $p(n_k = 0|r_k; r_{k-1}; \xi_k; s_j)$

Let $\gamma_j[\vec{r}(t)]$ be a position-dependent Poisson rate with $\vec{r}(t)$ some fixed path satisfying $\vec{r}(t_{k-1}) = \vec{r}_{k-1}$ and $\vec{r}(t_k) = \vec{r}_k$. Let $p_-[\vec{r}(t)]$ be the probability that no photon is emitted along the path $\vec{r}(t)$. From an elementary probability calculation, we have

$$p_-[\vec{r}(t)] = \exp\left(-\int_0^t dt'\, \gamma_j[\vec{r}(t')]\right) \tag{A.14}$$

where the notation is to indicate that p_- is a functional of the path $\vec{r}(t)$.

$p(n_k = 0|r_k; r_{k-1}; \xi_k; s_j)$ is the average of the functional over all (scalar) paths $r(t)$ with $r(0) = r_{k-1}$ and $r(t) = r_k$. Since the sample paths are generated by Brownian motion, we take the Wiener measure $d\mu_W[r(t)]$ with diffusion coefficient D_j as our probability measure on this function space [119]. We may now write $p(n_k = 0|r_k; r_{k-1}; \xi_k; s_j)$ as a path integral over the class $\mathcal{F}$ of all continuous functions from (r_{k-1}, t_{k-1}) to (r_k, t_k):

$$\begin{aligned} p(n_k = 0|r_k; r_{k-1}; \xi_k; s_j) &= \frac{\int_{\mathcal{F}} d\mu_W[r(t)]\, p_-(t|r(t))}{\int_{\mathcal{F}} d\mu_W[r(t)]} \\ &= \frac{1}{p(r_k|r_{k-1}; \Delta t_k; s_j)} \int_{\mathcal{F}} d\mu_W[r(t)] \exp\left(-\int_0^t \gamma_j[r(t')]dt'\right). \end{aligned} \tag{A.15}$$

If we now write the instantaneous rate function as

$$\gamma_j[r(t)] = \tilde{\gamma} + \left(\gamma_j[r(t)] - \tilde{\gamma}\right) \tag{A.16}$$

where

$$\tilde{\gamma} = \frac{1}{2}\left(\gamma_j[r_k] + \gamma_j[r_{k-1}]\right), \tag{A.17}$$

we may make a perturbation expansion of Eq. A.15 around the constant potential $\tilde{\gamma}$.

For the constant rate $\tilde{\gamma}$, we have

$$\int_{\mathcal{F}} d\mu_W[r(t)] \exp\left[-\int_{t_{k-1}}^{t_k} \tilde{\gamma} dt'\right] = p(r_k|r_{k-1}; \Delta t_k; s_j) \exp\left[-\tilde{\gamma}\Delta t_k\right]. \tag{A.18}$$

Including the first order contribution from the perturbing term $\gamma_j(r) - \tilde{\gamma}$ gives the next lowest-order correction

$$\begin{aligned} p(n_k = 0|r_k; r_{k-1}; \xi_k; s_j) &= \exp\left[-\tilde{\gamma}\Delta t_k\right] - \frac{2\pi}{p(r_k|r_{k-1}; \Delta t_k; s_j)} \\ &\times \int_{t_k}^{t_{k-1}} dt \int dr\, r p(r_k|r; t_k - t; s_j) \left[\gamma_j(r) - \tilde{\gamma}\right] p(r|r_{k-1}; t - t_{k-1}; s_j) + \cdots \end{aligned} \tag{A.19}$$

Appendix B

Statistical limits to dilute concentration estimation

B.1 Introduction

Chemical concentration is perhaps the most fundamental measurement one can make in the analysis of an arbitrary solution sample. With the recent success of single-molecule detection (SMD) techniques in measuring chemical reaction kinetics in dilute solutions and biological systems, the value of the ability to measure dilute concentrations has been made quite apparent[105, 120]. Simultaneously, the recent push to miniaturize the analytical apparatus required to make increasingly intricate chemical and biological measurements into monolithic lab-on-a-chip devices has introduced the requirement that concentration measurements be made accurately and quickly on a microscopic scale, using only a very small volume of sample[121, 122]. Single-molecule detection techniques have already been implemented for this purpose, and it is inevitable that their use will grow far more widespread as miniaturization continues due to their intrinsic ability to satisfy the requirements of these devices[123, 124]. However, SMD techniques, by design, collect information from only small numbers of discrete particles in any finite detection time, and this has statistical consequences: the value of the measured concentration can only be considered an *estimate*, the accuracy of which improves over time in a mathematically well-understood way. In order to make analytical measurements on increasingly dilute or increasingly small samples, it will be crucial to understand how the concentration estimation technique chosen will perform as a function of time. If concentration fluctuations occur on a time-scale that is too fast, or if an application requires a certain degree of accuracy in too short a time period, a particular technique may be inappropriate — in fact, it is possible that no technique is capable of reaching particular performance goals.

The traditional approach to spectroscopic concentration measurement is to measure some property of the sample that is proportional to the concentration, such as absorbance or fluorescence, within some region of the sample. A drawback associated with this approach is that at very low concentrations, or with very small detection regions, the concentration within the detection region exhibits fluctuations that grow increasingly large relative to the mean concentration, as illustrated in Fig. B.1. This fact is exploited by confocal Fluorescence Correlation Spectroscopy (FCS), which estimates concentration based on the variance of these fluctuations[26]. In higher-concentration regimes, smaller relative fluctuations about the mean lead to faster convergence of concentration estimates, such that temporal convergence properties are not important. In the single-molecule regime, however, large relative fluctuations lead to large uncertainty in concentration estimates made using averaging times that are too short. Unfortunately, the statistics of concentration fluctuations are very complicated because of their diffusive nature[125] such that characterization of the convergence properties of methods that rely on averaging the detection-region occupancy is quite difficult, and little work has been done in this area.

In this paper we address the problem of concentration estimation in *relatively* dilute samples, by which we mean that the average number of analyte particles within the detection region is much smaller than 1. At low relative concentrations it is possible to resolve the discrete entry times of individual analyte particles into the detection region, which allows for an alternative and fundamentally different approach to concentration estimation. The rate at which particles are detected is fully determined by the concentration, the geometry of the detection region and the diffusive properties of the particles. However, once detected, a particle's dynamics within the detection region are determined strictly by the geometry of the region and by its diffusive properties, assuming that the analyte particles do not interact. The concentration plays no role in determining the motion of the particles or the time spent by each particle within the detection region, so that the rate of individual particle detections is the only feature of the detected signal containing concentration-relevant information. We may discard nearly the entire signal, using only the time intervals between detections of individual particles to make statistically complete concentration estimates.

In the first section of this paper, we derive a probabilistic model for the time-spacing between detections of individual analyte particles. The exactness and simplicity of this model allow us, in the next section, to compute a limit for the convergence rate of a concentration estimate that is independent of the estimation technique, and hence to determine a lower bound for the kinetic time-scales that may ever be probed experimentally in dilute solutions.

We continue by deriving a concentration estimator using just the particle detection times extracted from an SMD signal. Based upon such a reduced signal, this technique is easily characterized in terms of its convergence, allowing us to explicitly show that its accuracy over time asymptotically approaches optimality. We formulate this estimator as a time-domain tech-

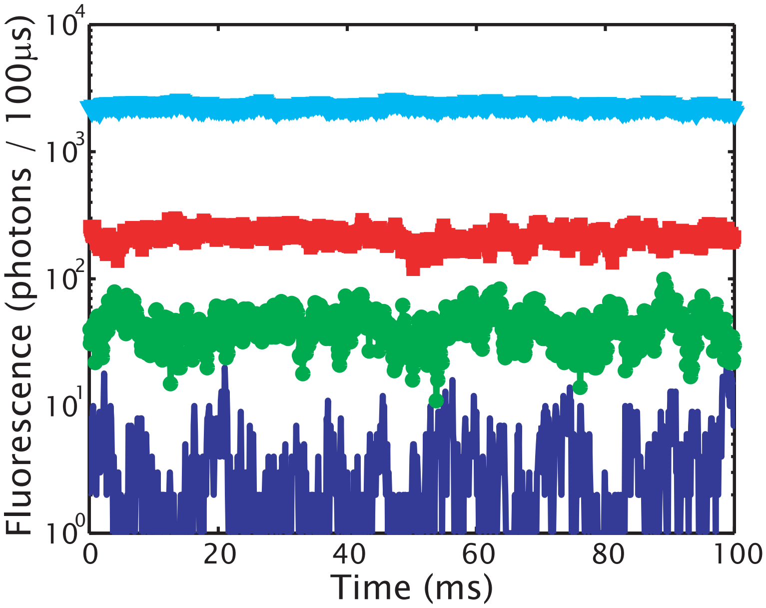


Figure B.1: Simulation of the concentration fluctuation statistics of several concentration regimes. Ensembles of 100ms Brownian motion trajectories were generated at relative concentrations of 0.1, 2, 10 and 100. Fluorescence photons were generated and the resulting fluorescence rates are shown.

nique, which is unique among SMD methods for concentration estimation because no averaging is required over multiple signal bursts. It is therefore naturally suited for use with methods that analyze individual signal bursts for independently monitoring the concentrations of the individual components of a sample.

In the final sections of this paper, we discuss the implementation of our technique experimentally and we present the results of detailed numerical simulations that illustrate its use in several realistic experiments. The highlight of these sections is the combined application of a Bayesian burst analysis method[16] with our concentration estimator to simultaneously monitor the concentrations both species in a binary solution.

B.2 Statistics of exact particle detection

In this section we derive a model for the concentration-dependent times between the detection of individual analyte particles when the distribution of particles in the sample is known and at equilibrium. We believe that the majority of cases of experimental interest will fit within these specifications — in any closed volume the distribution will be uniform, and in a microfluidic device the sample geometry is so tightly confined that the analyte distribution should be calculable to within the necessary accuracy. By assuming that the distribution of particles is in equilibrium, we are technically limiting ourselves to cases in which the sample concentration

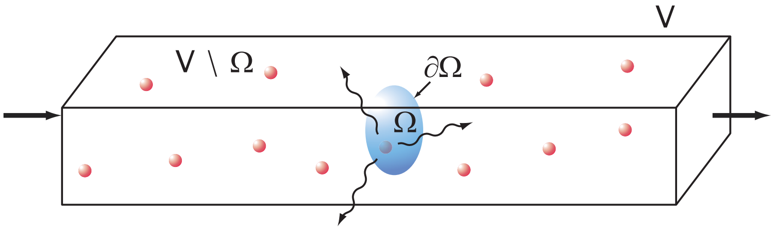


Figure B.2: Illustration of the regions of the sample that we describe in the text, in the example of a microfluidic channel. V is the entire sample, Ω is the detection region, $V \setminus \Omega$ is the region in the sample outside of Ω, and $\partial\Omega$ is the boundary of Ω. Red spheres represent analyte particles. Wavy arrows indicate fluorescence emission. Large arrows outside of V indicate the direction of fluid flow.

is held constant. In practice this will only be an absolute requirement over a time period sufficiently long to make accurate concentration estimates; we avoid unnecessary complications by leaving such fluctuations out of our model, and we discuss our approach to dealing with slowly fluctuating concentrations in later sections. Rapidly fluctuating concentrations cannot be ignored from the model in this way; rarely, however, will it be possible to accurately monitor such fluctuations experimentally in dilute solutions.

We assume that in our experimental apparatus there is a region Ω within the overall sample volume V that we monitor for the presence of analyte particles (see Fig. B.2 for illustration). In single-molecule fluorescence experiments, for example, Ω is the region defined by the focus of the excitation laser and the fluorescence-collection optics.

Given that the equilibrium spatial distribution $c(x)$ of the analyte particles may be nonuniform — microfluidic devices containing regions in which two sample streams are mixed are a good example — we must define precisely what we mean by the concentration of the sample. We define the normalized spatial density $\rho(x)$ and the concentration C such that

$$\int_\Omega \rho(x)dx = |\Omega|, \tag{B.1}$$

where $|\Omega|$ is the volume of Ω, and

$$C\rho(x) = c(x), \tag{B.2}$$

which is consistent with standard notions of concentration and rigorously accounts for inhomogeneity. Throughout the paper, we will frequently refer to concentration in units relative to $|\Omega|$, such that numerical concentrations will be dimensionless and $|\Omega| = 1$. Under this convention, C is the average number of particles within Ω over time. This is done to underscore the importance of relative concentrations over absolute concentrations in our presentation.

B.2.1 Detection-region diffusive influx dynamics

We begin by deriving an expression for the influx rate of analyte particles into an arbitrary detection region Ω. We assume that the motion of analyte particles obeys the convection-diffusion equation

$$\frac{\partial}{\partial t}\rho(x,t) = \nabla \cdot (u(x)\rho(x,t)) + D\nabla^2\rho(x,t), \tag{B.3}$$

where D is the diffusion coefficient of the particles and $u(x)$ is an arbitrary fluid flow field. In order to solve for the rate at which particles enter Ω, we follow a procedure originally developed by Smoluchowski to study irreversible chemical kinetics. According to this model, the rate at which particles reach the surface $\partial\Omega$ of the region Ω can be found by solving Eq. B.3 over $V \setminus \Omega$ (the region of V excluding Ω) with the absorbing boundary condition $\rho(x,t) = 0$ on $\partial\Omega$ and appropriate initial conditions[8, 126, 127]. Since $\rho(x,t)$ is in equilibrium with V, all of the net flux of analyte out of $V \setminus \Omega$ is through $\partial\Omega$, such that the rate $\gamma(t)$ with which particles enter Ω is given by the change in the number of analyte particles in $V \setminus \Omega$,

$$\gamma(t) = -C\frac{\partial}{\partial t}\int_{V\setminus\Omega} dx\rho(x,t). \tag{B.4}$$

Combining Eqs. B.3 and B.4 and applying Green's theorem, we have

$$\gamma(t) = -CD\int_{\partial\Omega} \nabla\rho(x,t) \cdot dA, \tag{B.5}$$

where the integrand is the familiar diffusive flux into Ω. Note that the convective term has vanished because $\rho(x,t)$ is zero on $\partial\Omega$, but is incorporated into the integral through its effect on $\nabla\rho(x,t)$.

The reduction of the detection-region influx statistics from a multi-dimensional PDE model to a scalar function of time is a significant result, particularly in microfluidic applications. It allows us to reduce arbitrarily complicated sample and detection geometries combined with diffusion statistics into a single function that exactly characterizes the concentration-relevant statistics in an experiment. Furthermore, that $\gamma(t)$ depends linearly on C is a noteworthy point. It allows us to define the concentration-independent rate function

$$\tilde{\gamma}(t) = \frac{1}{C}\gamma(t) \tag{B.6}$$

that is an intrinsic characteristic of the sample and detection geometries. This separation implies that for a given geometry and sample species, we must determine $\gamma(t)$ only once in order to fully specify the detection-region influx statistics at any concentration.

In practical terms, computing $\tilde{\gamma}(t)$ is a nontrivial task for all but the simplest sample geometries. It will sometimes be the case, however, that the steady-state value of $\tilde{y}(t)$ alone is sufficient to approximately characterize the influx statistics. In particular, this will be true when $\tilde{y}(t)$ decays to steady-state quickly relative to $C\tilde{y}(t)$, the rate at which particles are detected. Such rapid decay is a characteristic of experiments in which convective flow drives the motion of particles into Ω, causing $\tilde{y}(t)$ to decay rapidly, and is also found in free solution at very low concentrations. It results in only very few particles entering Ω before the difference between $\tilde{y}(t)$ and $\tilde{y}(\infty)$ is negligible. A concentration estimate made using a fairly large number of detections will be trivially influenced by the time dependence of these few; therefore we may solve for only the steady-state influx rate, which is a much simpler computational task due to the time-independence of the problem.

B.2.2 Particle-detection statistics

Since particle detection is a discrete process, the rate $C\tilde{y}(t)$ at which we detect new particles is only meaningful in a probabilistic sense. In this section we complete our model by deriving a probability density over the time-spacing between discrete detection events.

Imperfections in our ability to extract detection times from an SMD signal will affect the statistics we observe. Here we will ignore the exact sources and consequences of these imperfections, but assume that we may still define a rate of particle detection $j(t)$ that is distinct from but related to the rate of diffusive influx $\tilde{y}(t)$. If we assume that detection is perfect — that all particles entering Ω are detected immediately — then $j(t) = \tilde{y}(t)$. We will restrict ourselves to this approximation until a later section, when we provide more details on the types of inaccuracies inherent to most detection methods.

For the purpose of derivation, we define the conditional probability $P(\tau_k|t_{k-1})$ that the k^{th} particle in a stream of particles is detected at some time $T > \tau_k + t_{k-1}$ *given* that the $(k-1)^{th}$ particle was detected at time t_{k-1}. We may write

$$P(\tau_k + d\tau_k|t_{k-1}) = P(\tau_k|t_{k-1})\left(1 - Cj(\tau_k + t_{k-1})d\tau_k\right) \tag{B.7}$$

since the probability that no particle is detected in the infinitesimal interval $d\tau_k$ is $1 - Cj(\tau_k + t_{k-1})d\tau_k$. Rearranging Eq. B.7 and taking the limit $d\tau_k \to 0$, we have the differential equation

$$\frac{\partial}{\partial \tau_k} P(\tau_k|t_{k-1}) = -Cj(\tau_k + t_{k-1})P(\tau_k|t_{k-1}), \tag{B.8}$$

the solution to which is the exponential decay

$$P(\tau_k|t_{k-1}) = \exp\left\{-C\int_0^{\tau_k} dt'\, j(t' + t_{k-1})\right\}. \tag{B.9}$$

Finally, we note that $1 - P(\tau_k|t_{k-1})$ is, by definition, the cumulative probability function corresponding to the probability density $p(\tau_k|t_{k-1})$ over the waiting time τ_k. Thus we have

$$p(\tau_k|t_{k-1}) = -\frac{\partial}{\partial \tau_k} P(\tau_k|t_{k-1}) = Cj(\tau_k + t_{k-1}) \exp\left\{-C \int_0^{\tau_k} dt' j(t' + t_{k-1})\right\}. \tag{B.10}$$

The result that we will require in later sections is not simply the probability of a single waiting time, but rather the joint probability of the set $\xi = \{\tau_1, \ldots, \tau_k\}$ of sequential waiting times over the course of an experiment. It is simple to show that this joint probability is given by

$$p(\xi|C) = C^k j(\tau_1) \cdots j\left(\sum_{i=1}^{k} \tau_i\right) \exp\left[-C\zeta\left(\sum_{i=1}^{k} \tau_i\right)\right], \tag{B.11}$$

where we have made the substitution

$$\zeta(t) = \int_0^t dt' j(t'). \tag{B.12}$$

Equation B.11 is the general probabilistic model that we will use in all forthcoming derivations in this paper.

B.3 Concentration estimation

In this section we restrict ourselves to purely statistical calculations using the particle detection model derived in the previous section. We begin by deriving a performance limit for the convergence of concentration estimates assuming only that the concentration estimation technique used is unbiased. Next we derive and characterize a simple time-domain concentration estimator, the convergence of which asymptotically approaches this limit. The calculations in this section are independent of the particle-detection technique in the sense that the only assumption we make is that the model for the particle-detection statistics, namely that the detection rate is linear in C, is correct.

B.3.1 Universal limit to estimator performance

It is a critically important point that the statistics of diffusive influx into Ω are a complete model of the concentration-relevant information present in the signal that we detect in an SMD experiment. The occupancy of Ω at any time is determined by the competing rates of particles diffusing into and out of Ω. However, once a particle has entered Ω, its exit time depends only on the geometry of Ω and on its diffusive dynamics; while the spatial density $\rho(x)$ may affect the exit statistics by determining where particles are most likely to enter Ω, the scalar concentration C can only affect the exit times if the diffusive dynamics of a single particle depend on the presence of other particles. If we assume that the analyte particles do not interact, we can safely ignore the detection-region outflux statistics without compromising any concentration information in the detected signal. The result is that any statistical properties we derive using our influx statistics model may be broadly applied to *any* concentration-estimation technique, including detection-region occupancy-averaging methods and FCS, because those methods simply cannot incorporate any additional information into their estimates.

Suppose we detect a set of detection-region entry times $\xi = \{\tau_1, \ldots, \tau_k\}$ under experimental conditions satisfying the assumptions of our detection statistics model. We define an estimator $\Upsilon(\xi)$ as a function of those entry times that we require to be unbiased, such that

$$E_C[\Upsilon(\xi)] = C, \tag{B.13}$$

where

$$E_C\left[\cdot\right] = \int_0^\infty d\tau_1 \cdots \int_0^\infty d\tau_k p(\xi|C)\left[\cdot\right] \tag{B.14}$$

denotes the expected value functional over all possible sets of entry times for a fixed C, and $p(\xi|C)$ is given by Eq. B.11.

We will use the variance of Υ,

$$\mathrm{Var}_C[\Upsilon(\xi)] = E_C\left[(\Upsilon(\xi) - C)^2\right], \tag{B.15}$$

as a standard measure of its performance — smaller variance indicates that $\Upsilon(\xi)$ is distributed more closely to the actual value of C, indicating faster convergence of the estimator. A fundamental property of all unbiased estimators is that their variance satisfies a positive lower bound for any finite measurement record. This is simply a consequence of the limited amount of statistical information contained in a finite sample. A concentration estimator $\Upsilon(\xi)$ based on particle detection statistics satisfies the assumptions of the Cramér-Rao inequality, which gives us the lower bound on $\mathrm{Var}_C[\Upsilon(\xi)]$[85],

$$\mathrm{Var}_C[\Upsilon(\xi)] \geq \frac{1}{E_C\left[\left(\frac{\partial}{\partial C}\log p(\xi|C)\right)^2\right]}. \tag{B.16}$$

Substituting Eq. B.11 into Eq. B.16, the Cramér-Rao inequality evaluates to

$$\mathrm{Var}_C\left[\Upsilon(\xi)\right] \geq \frac{C^2}{k}, \tag{B.17}$$

representing a performance limit for any unbiased concentration estimation method.

An unbiased estimator that achieves equality in Eq. B.17 is referred to as *optimal*. Almost universally, the goal in estimation problems is to find an optimal estimator; however, one does not always exist. In fact, by applying the attainment corollary to the Cramér-Rao inequality to our detection statistics model it can be shown that no optimal concentration estimator exists[85]. We conclude, therefore, that regardless of the estimation technique used, the concentration estimate will always converge with variance strictly greater than C^2/k.

We note that nothing about our derivation of Eq. B.17 requires that detection is done via SMD. In particular, Eq. B.17 applies even for large-scale concentration measurement devices such as spectrophotometers or fluorometers. However, such devices monitor large regions of relatively high-concentration samples so that $C\tilde{y}(t)$ is very large. As a result, k becomes large very quickly so that C^2/k is negligible and we can consider measurements to have fully converged in very short times. In SMD, $C\tilde{y}(t)$ is intentionally restricted for the sake of single-molecule resolution, and as a result C^2/k is nontrivial.

B.3.2 Near-optimal time-domain concentration estimator

One procedure for finding an estimator for a particular quantity is to compute the maximum likelihood estimator $\Upsilon_{ML}(\xi)$, defined such that[85]

$$p(\xi|\Upsilon_{ML}(\xi)) = \max_C p(\xi|C). \tag{B.18}$$

The maximum likelihood estimator for C is given by

$$\Upsilon_{ML}(\xi) = \frac{k}{\zeta\left(\sum_{i=1}^{k} \tau_i\right)}, \tag{B.19}$$

for which we compute the expected value

$$E_C\left[\Upsilon_{ML}(\xi)\right] = C\frac{k}{k-1}, \tag{B.20}$$

indicating that $\Upsilon_{ML}(\xi)$ is biased. It is clear from Eq. B.20, however, that the modification of $\Upsilon_{ML}(\xi)$ given by

$$\tilde{\Upsilon}(\xi) \equiv \frac{k-1}{k}\Upsilon_{ML}(\xi) = \frac{k-1}{\zeta\left(\sum_{i=1}^{k} \tau_i\right)} \tag{B.21}$$

is unbiased for $k > 1$; additionally, since $(k-1)/k < 1$, we know that the variance of $\tilde{\Upsilon}$ will be strictly smaller than the variance of Υ_{ML}. Therefore $\tilde{\Upsilon}$ is an unequivocally better estimator than

Υ_{ML}, so $\tilde{\Upsilon}$ is the estimator that we will characterize and use. The variance of $\tilde{\Upsilon}$ is given by

$$\mathrm{Var}_C\left[\tilde{\Upsilon}(\xi)\right] = \frac{C^2}{k-2}, \tag{B.22}$$

which compares favorably to the optimal variance of Eq. B.17 and approaches optimal in the limit $k \to \infty$. Because $\tilde{\Upsilon}$ is unbiased, we can use Eq. B.22 to give uncertainty bounds in our estimates of the standard form $\tilde{y}(\xi) \pm 2\sqrt{\tilde{y}(\xi)^2/(k-2)}$, approximately corresponding to a 95% confidence interval.

B.3.3 Time-domain variance and concentration fluctuations

The optimal variance given by Eq. B.17 and the variance of $\tilde{\Upsilon}$ given in Eq. B.22 are both expressed as functions k, but in a fixed amount of time this number depends on the analyte concentration such that these expressions do not clearly represent the convergence of the estimates over time. To clarify this time dependence, we compute the expected value of the variance in Eq. B.22 as an explicit function of time:

$$E_{t,C}\left[\mathrm{Var}_C\left[\tilde{\Upsilon}(\xi_t)\right]\right] = \sum_{k_t=3}^{\infty} \frac{C^2}{k-2} p(k_t), \tag{B.23}$$

where ξ_t is the measurement record up to time t, k_t is the number of particles detected up to time t, and $p(k_t)$ is the probability of detecting k_t particles in time t given by the Poisson distribution,

$$p(k_t) = \frac{1}{k_t!} \exp\left[-C\zeta(t)\right]\left[C\zeta(t)\right]^{k_t}. \tag{B.24}$$

Note that the sum in Eq. B.23 begins at $k_t = 3$ because Eq. B.22 is infinite for $k < 3$; Eq. B.23 is therefore a *conditional* expectation, in which it is assumed that at least 3 particles have been detected. Equation B.23 is invalid for very small t but becomes quite accurate very quickly as the probability that only 2 or fewer particles have been detected becomes vanishingly small for large t. While the summation in Eq. B.23 cannot be computed explicitly, it can be represented in terms of the generalized hypergeometric function F,

$$E_{t,C}\left[\mathrm{Var}_C\left[\tilde{\Upsilon}(\xi_t)\right]\right] = \frac{C^2}{6}\left[C\zeta(t)\right]^3 \exp\left[-C\zeta(t)\right] F\left[\begin{matrix} 1,1 \\ 2,4 \end{matrix}; C\zeta(t)\right], \tag{B.25}$$

which simplifies evaluation due to numerical methods for computing F.

Now that we have an expression for the variance over time of $\tilde{\Upsilon}$ at a fixed concentration, we can discuss the measurement of a dynamically-fluctuating concentration $C(t)$. In order to permit a time-dependent concentration within the framework we developed for computing the particle detection statistics, we require that $C(t)$ fluctuates sufficiently slowly that its time dependence does not affect the form of $\tilde{y}(t)$ — that is, the fluctuations are slow enough that they produce negligible changes in the equilibrium density $\rho_0(x)$ in V. This time-scale is set by the time required for concentration fluctuations to propagate throughout the sample by convective and diffusive motion. We must also require that $C(t)$ is approximately constant for a time sufficiently long to estimate it with a suitably small variance. In the single-molecule concentration regime, the latter will nearly always be the dominant time constraint.

Provided a $C(t)$ with sufficiently slow fluctuations, our estimation strategy is simple. We begin by choosing an acceptable uncertainty in the concentration estimate and using Eq. B.23 to approximate the estimation time required to attain it. We divide the detected signal into time intervals of this required time and estimate the approximately-constant C within each interval using $\tilde{\Upsilon}$. Finally, we construct $C(t)$ using these estimates. We present simulations in a later section that illustrate this strategy.

B.4 Experimental considerations

The results in the previous section rely on the assumption that the time-dependent detection rate of analyte particles scales linearly with concentration. We have already shown that this is true in the limit of perfect detection, when detection-region entry times can be determined exactly. In this section, we begin by discussing how realistic sources of detection inaccuracy do not violate this linear dependence. Next we discuss how the inability to distinguish individual particles gives rise to a problem of recurrent detections, in which one particle follows multiple sequential paths through the detection region and the resulting signal bursts are attributed to multiple distinct particles. We go on to discuss techniques for extracting detection times and rejecting recurrent detections from a single-molecule fluorescence signal. Finally we discuss the effects of multiple occupancy, in which a concentration-dependent fraction of particles entering Ω will go undetected because Ω was not empty upon their entry.

B.4.1 Localization inaccuracy

The inability to precisely localize individual particles in SMD experiments[128] has two primary consequences that are of concern to us. First, once a particle has entered Ω, some time will be required before we detect it and there is some probability that it will leave Ω before this time, escaping detection. Second, it is possible that particles following trajectories that come close to Ω without actually entering it will be detected as having entered.

We may model the first of these inaccuracies using the conditional probability $\eta(\tau|t)d\tau$

that a particle is detected at time $t + \tau$, *given* that it entered Ω at time t. Note that $\eta(\tau|t)$ is in general not normalized, because there is some nonzero probability that a particle will escape detection. If fluorescence is the method of detection, η has a sharp peak at $t = 0$ because the time-scale of fluorescence is much shorter than that of diffusion. As a result, we can make the approximation $\eta(\tau|t) \approx \delta(\tau)$. To allow for arbitrary detection methods, however, we do not restrict the form of $\eta(\tau|t)$.

The second inaccuracy may be modeled as a time-dependent rate of spurious detections $Cj_0(t)$, where the proportionality to C follows from its dependence on the concentration of analyte particles immediately outside Ω. Typically we will choose Ω in such a way that the spatial sensitivity of the detection method is high on $\partial\Omega$, so that $j_0(t)$ will be fairly small. As with η, however, we will not restrict the values of $j_0(t)$.

Using $C\tilde{y}(t)$ from Eq. B.6 as the rate that particles actually *enter* Ω we may define $Cj(t)$, the rate at which particles are *detected*, as

$$Cj(t) = Cj_0(t) + C\int_0^t dt'\eta(t - t'|t')\tilde{y}(t'). \tag{B.26}$$

In the same sense that $\tilde{y}(t)$ is a characteristic of the diffusive dynamics alone, $j(t)$ is a characteristic of the coupled diffusion and detection dynamics in a particular experiment. Its linearity for general η and j_0 imply that all of the results of the previous section will hold for any detection method subject to our assumptions here regarding localization inaccuracy.

B.4.2 Recurrent detections

The detection statistics that we use in this paper are based on a model in which particles are only counted the *first* time they enter Ω, because this is the only entry that gives us any concentration-relevant information. Unfortunately, we cannot easily distinguish between particles that have already been detected and those that have not. Thus, whatever technique we use to determine detection times must correct for the recurrent detections of single particles, in which an analyte particle is detected, exits the detection region, re-enters at some later time and is detected once again. If ignored, these events will drive the detected analyte influx rate much higher than the calculated rate for a given geometry and concentration, giving a falsely high concentration estimate.

Fortunately, the time-scale of recurrent detections is a result of the detection-region geometry alone and does not depend on the concentration of the sample. This time-scale tends to be very fast because a particle is very likely to re-enter Ω almost immediately after exiting, while re-entry becomes less likely over time as the probability that the particle has remained close to Ω decreases. Thus, in dilute samples we expect a separation between the time-scales

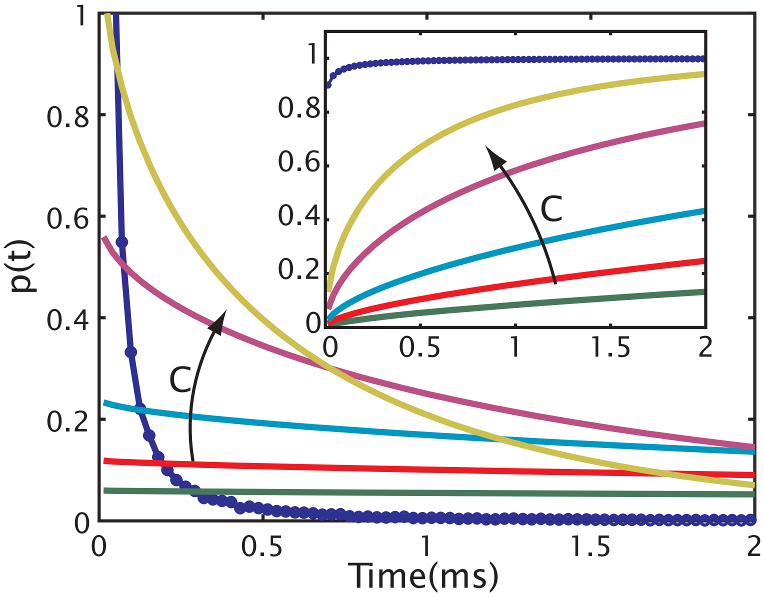


Figure B.3: Comparison of the time-scales of recurrent detections and first detections for freely diffusing particles into a spherical detection region with radius 0.5μm. The main plot compares the probability densities of recurrent detection times (blue curve with circle markers) with those of first detection times (curves without markers) for relative concentrations of 0.05, 0.1, 0.2, 0.5 and 1. The inset compares the cumulative probabilities corresponding to those probability densities. Arrows indicate the direction of increasing concentration. Recurrent detection statistics were generated by Monte-Carlo simulation, while the exact solution to Eq. B.3 was used for first detection.

of initial and recurrent detections. Figure B.3 shows the recurrent-detection time probability compared to the probability of first-entry times for particles freely diffusing into a spherical Ω in an unbounded V. At concentrations relevant to SMD, the time-scale separation is great enough that simply ignoring short detection times will reject nearly all recurrent detections while rejecting very few first detections.

B.4.3 Entry-time determination methods

In order to make concentration estimates using our detection-time estimator, we require the ability to accurately extract detection times from an SMD signal. Two tools have been employed to do this for fluorescence-based SMD while simultaneously identifying recurrent particle detections. The first of these is the Lee filter, in which the detected signal is smoothed such that fluorescence bursts from recurrent detections are merged while the fine structure of the bursts is preserved and may be used for subsequent analysis[129]. These merged signal bursts are identified by simply applying a fluorescence count threshold to the signal.

The second approach is to use the Bayesian filter that we have introduced, which computes a spatial probability density for the detected particle over time[16]. Applying a threshold to the probability that Ω is non-empty is an accurate way to determine entry times, and the spatial estimator's ability to monitor particles over some distance as they diffuse away from Ω provides a criterion for identifying recurrent detections. This method requires considerably more computational power to implement than the Lee filter, but it enables experiments in which each signal burst is analyzed individually, so that the concentrations of multiple species may be directly monitored simultaneously in a single sample. For simple experiments where little other than the concentration of a homogeneous solution is needed, the Lee filter is ideal because of its simplicity. The Bayesian filter is the natural technique for more complicated experiments, as it provides a rigorous framework for simultaneous detection and analysis of signal bursts.

B.4.4 Bias due to multiple occupancy

Single-molecule detection experiments are typically designed under the assumption that the concentration of the sample is low enough that it is very unlikely that more than one particle will be within the detection region at a given time. The probability of detecting two or more particles is always nonzero, however, and we must consider how this affects our concentration measurements. The number of particles inside Ω at any time has a Poisson distribution, so the probability that Ω is empty is $\exp(-C|\Omega|)$. We will only be able to detect particles that enter Ω while it is empty, so that our estimator will converge to an effective concentration

$$C_{eff} \equiv E_C\left[\tilde{\Upsilon}(\xi)\exp(-C|\Omega|)\right] = C\exp(-C|\Omega|) \tag{B.27}$$

that is biased by the exponential term. When the relative concentration $C|\Omega|$ is small $C_{eff} \approx C$, so that this bias is negligible; however, it cannot be ignored in the higher range of SMD-relevant relative concentrations.

In order to recover the sample concentration from the estimated effective concentration, we define the corrected concentration estimator $\tilde{\Upsilon}_{corr}$ by inverting Eq. B.27:

$$\tilde{\Upsilon}_{corr}(\xi) = -\frac{1}{|\Omega|}W\left[-|\Omega|\tilde{\Upsilon}(\xi)\right], \tag{B.28}$$

where W is the Lambert W function[130]. We know that in the limit $k \to \infty$, $\tilde{\Upsilon}$ converges in probability to C_{eff}, and since W is continuous we know that $\tilde{\Upsilon}_{corr}$ must converge to C[131]. Thus we know that $\tilde{\Upsilon}_{corr}$ will be unbiased asymptotically, but we require knowledge of its mean and variance for finite k to determine how many particles we must detect in order to make acceptably accurate estimates.

The functional form of W makes exact computation of $E_C\left[\tilde{\Upsilon}_{corr}\right]$ and $\mathrm{Var}_C\left[\tilde{\Upsilon}_{corr}\right]$ difficult. Instead, we computed these statistics by numerical integration in MATLAB (Mathworks, Inc.,

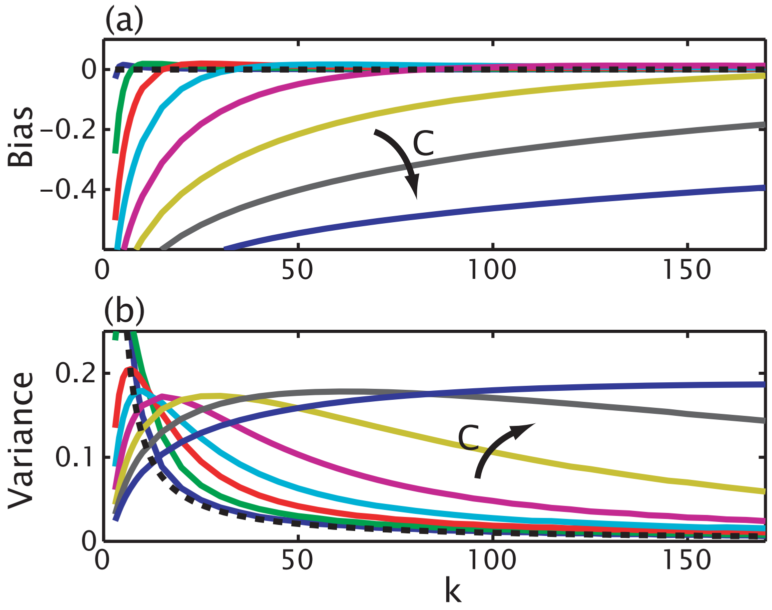


Figure B.4: Statistical properties of $\tilde{\Upsilon}_{corr}$ at relative concentrations of 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65 and 0.75. The plot in (a) shows the bias of $\tilde{\Upsilon}_{corr}$ relative to C, defined as $C^{-1}E_C\left[\tilde{\Upsilon}_{corr}(\xi) - C\right]$, and (b) shows the variance of $\tilde{\Upsilon}_{corr}$ relative to C^2, defined as $C^{-2}\,\mathrm{Var}_C\left[\tilde{\Upsilon}_{corr}\right]$ (solid curves), and that of $\tilde{\Upsilon}$, given by $(k-2)^{-1}$ (black dashed curve). In both plots, arrows indicate the direction of increasing concentration.

Natick, MA) over a range of relative concentrations that are relevant for SMD and a over range of values for k. These results are shown in Fig. B.4. The first plot is the relative estimator bias, showing the error that we expect in our estimates as a percentage of the sample concentration C. The second plot shows the variance of $\tilde{\Upsilon}_{corr}$ at several concentrations, compared to the variance of $\tilde{\Upsilon}$ from Eq. B.22.

Together, the plots in Fig. B.4 express our requirement for low relative concentrations in quantitative terms. For relative concentrations less than 0.25, $\tilde{\Upsilon}_{corr}$ has a negative bias only for small k, such that after approximately 50 particles are detected the estimator is effectively unbiased. Furthermore, after 50 particles are detected the variance of $\tilde{\Upsilon}_{corr}$ is very close to that of $\tilde{\Upsilon}$ throughout this concentration range. We should reasonably expect that even with perfect detection we would require this many or more detections to achieve an acceptable estimator variance, so the detection-induced bias will need to be paid only slight attention at these low concentrations. At higher concentrations, however, $\tilde{\Upsilon}_{corr}$ quickly becomes very inaccurate for any reasonable estimation time. Therefore, when we specify that we require a low relative concentration, we can take this to mean one that is not much higher than 0.25.

Of course, the results in this section would be irrelevant if we had the ability to determine with infinite resolution the number of particles within the detection region at any time. While some techniques could be adapted to resolve multiple particles within the detection region it is difficult to do this accurately for even just a few particles and nearly impossible for larger numbers. Therefore, while we may be able to extend the utility of our technique to higher concentrations, there will always be a limit to how high that concentration may be.

B.5 Simulations

In this section we present the results of numerical simulations to illustrate the key points from the derivations in the previous sections. All simulation codes were written in C and compiled and run in Linux, and are available from the authors upon request. Uniform pseudo-random numbers were generated using the publicly-available Mersenne Twister code[132], and standard techniques were used to transform these to random numbers of the appropriate distributions[133]. All data analysis was done with MATLAB and C code, also available from the authors.

Ensembles of three-dimensional Brownian motion trajectories were generated at fixed or fluctuating concentrations, specified in the text. Simulated particles had a diffusion coefficient $D = 100\mu\text{m}^2\text{s}^{-1}$, unless specified otherwise. Their positions were updated after every time step $\Delta t = 100\text{ns}$ with independent random Gaussian-distributed steps in each Cartesian direction with variance $2D\Delta t$ and mean $u(r)\Delta t$ determined by the convective flow field, if applicable. Simulations were limited to bounded regions interfaced to constant-concentration reservoirs in a manner similar to that in [134] but in 3 dimensions and in the presence of convection. Free diffusion simulations were performed within a sphere of radius 3 μm and channel simulations, within a 20μm cylinder with radius 1μm and reflecting boundary conditions imposed on the radial boundary. The particles in the channel were driven by the laminar flow field $u(r) = \alpha(R^2 - r^2)\mathbf{e}_z$, where R is the outer radius of the channel, α was adjusted to yield a peak velocity of 5mm/s, and $\mathbf{e}_z$ is the unit vector parallel to the channel. These geometries were chosen in particular for their simplicity and their experimentally relevant size scales.

B.5.1 Concentration estimation with exact particle detection

Our first set of simulations is meant to illustrate the performance of $\tilde{Y}$ under ideal conditions — when our knowledge of the detection-region entry times is exact.

We begin with simulations of free diffusion with a 0.5μm-radius spherical detection region. The symmetry of Ω allowed us to compute $\tilde{y}(t)$ exactly, given by

$$\tilde{y}(t) = 4\pi DR\left[1 + R\left(\pi Dt\right)^{-1/2}\right]. \tag{B.29}$$

The concentration of the sample relative to Ω was fixed at 0.1, or approximately 317pM in this

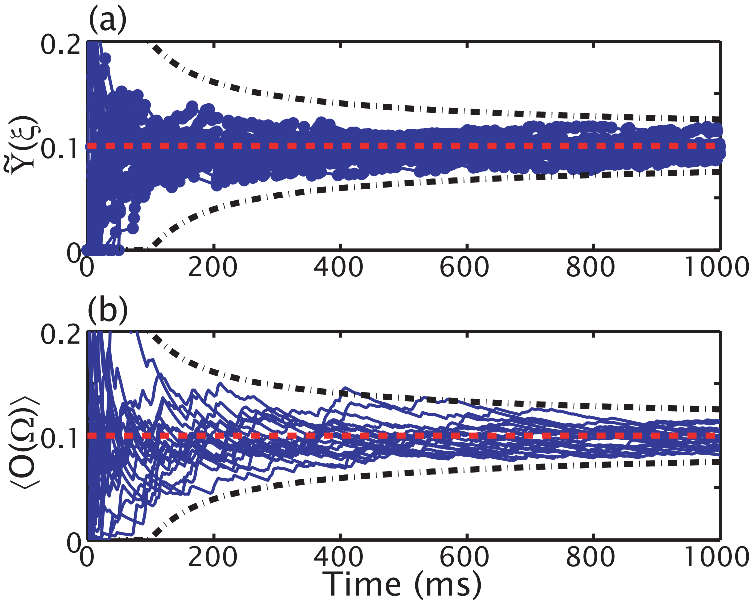


Figure B.5: Concentration estimation for free unbiased diffusion. Twenty $1s$ ensemble trajectories were simulated with a relative concentration of 0.10 (red dashed line). Concentration estimates (solid blue curves) were produced by (a) $\tilde{\Upsilon}$ and (b) $\langle O(\Omega)\rangle$, the time-averaged occupancy of Ω. Exact influx times were used in (a), and exact occupancy information was used in (b). Markers in (a) indicate particle-detection times. 2σ error bounds (black dot-dash curves) were computed using Eq. B.25.

geometry. Upon initial entry into Ω, particles were marked in order to eliminate recurrent detection. To properly preserve the spatial density of undetected particles within the simulation volume, we labeled newly-injected particles as having been already detected with a probability computed using the exact time-dependent solution to Eq. B.3. These correspond to particles that were detected at some time, exited the simulation volume, and then re-entered the simulation volume at a later time. Using the exact particle locations, we computed $\tilde{\Upsilon}$ over time and compared it to the averaged occupancy of Ω over time. The results are shown in Fig. B.5. These results agree very well with the predictions of our detection statistics model. In addition, they show that the occupancy-averaging estimator converges with roughly the same variance as that which we computed in Eq. B.22, agreeing with our assertions about the generality of our convergence results and indicating that occupancy-averaging is a nearly-optimal concentration estimation method.

Next, we simulated twenty independent 250ms ensembles of trajectories of particles diffusing within a cylindrical microchannel. We defined Ω as an ellipsoid with two 0.5μm semiaxes and a third semiaxis perpendicular to the flow with length 0.8μm; Fig. B.2 shows an illustration of this geometry. The relative concentration of the simulation was fixed at 0.05, corresponding to 99pM. Given the incompatible symmetries of V and Ω, it is not possible to analytically

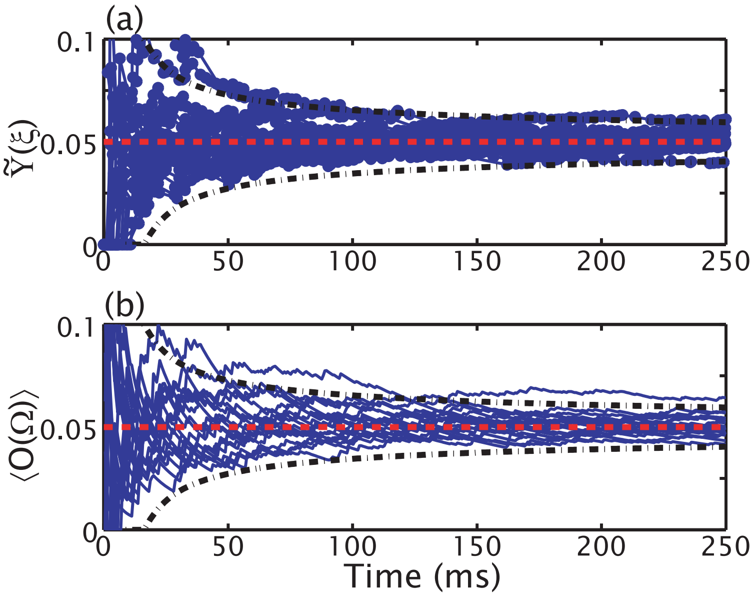


Figure B.6: Concentration estimation within a micro-channel. Twenty 1s ensemble trajectories were simulated with a relative concentration of 0.05 (red dashed line). Concentration estimates (solid blue curves) were produced by (a) $\tilde{\Upsilon}$ and (b) $\langle O(\Omega)\rangle$, the time-averaged occupancy of Ω. Exact influx times were used in (a), and exact occupancy information was used in (b). Markers in (a) indicate particle-detection times. 2σ error bounds (black dot-dash curves) were computed using Eq. B.25.

compute $\tilde{\gamma}(t)$. Instead, we determined $\tilde{\gamma}(\infty) \approx 9\text{ms}^{-1}$ by simulation and used this steady-state approximation to $\tilde{\gamma}(t)$ to make concentration estimates, as before, using exact occupancy and detection time statistics. The results are shown in Fig. B.6. These results highlight the generality of the detection-time concentration estimation technique by illustrating its accuracy for bounded, biased diffusion, and further support the observation that the variance bound we computed for the detection-time estimator approximates that of occupancy-averaging estimators.

B.5.2 Concentration estimation via fluorescence detection

In real single-molecule experiments we can never be certain of the actual positions of analyte particles, so the results just presented represent the best possible results we could ever expect to achieve experimentally. In order to evaluate more realistic scenarios, we performed simulations in which diffusing particles emitted fluorescence photons and concentration estimates were made by smoothing the fluorescence signal with a Lee filter and applying a count threshold to determine detection times. It is difficult to precisely define Ω in this case due to the smooth leading edge and irregular geometry of the laser focal region. As a result, in these and later simulations we fix the absolute concentrations but provide approximations for Ω and the relative concentrations.

The fluorescence rate of each particle in the sample was computed at each time step according to the spatially-modulated rate $\Gamma(r) = \Gamma_L(r)\Gamma_r / (\Gamma_L(r) + \Gamma_r) + \Gamma_B$ corresponding to a 2-level saturating-emitter model where Γ_r is the relaxation rate of the fluorophore in its excited state, Γ_B is the background fluorescence rate, and $\Gamma_L(r)$ is the position-dependent laser excitation rate in 3 dimensions. In the simulations we used Γ_r =1MHz and Γ_B =5kHz, and for Γ_L we used the Gaussian-Lorentzian beam profile

$$\Gamma_L(x, y, z) = \frac{\Gamma_L(0)}{w(z)^2} \exp\left[\frac{-2(x^2 + y^2)}{w(z)^2}\right], \tag{B.30}$$

where $\Gamma_L(0)$ =600kHz was the peak excitation rate and the beam waist $w(z)$ is given by $w(z)^2 = w_0^2\left[1 + (\lambda z/(\pi w_0^2))^2\right]$ with excitation wavelength λ =532nm and focused waist $w_0 = 0.5\mu$m. The fluorescence rates of all particles in the sample were added and the resulting total rate was used to generate exponentially-distributed waiting times between detected photons at each time step. The resulting signal was down-sampled into a sequence of photon counts within time windows of duration 50μs. An example of such a signal is shown in Fig. B.7. The signals were then smoothed using a Lee filter with a width of 350μs and a smoothing parameter $\sigma = 4$[129]. The recurrence-rejection properties of the Lee filter are evident in the inset of Fig. B.7, where fast fluctuations causing the raw signal to climb above and then dip below the threshold value — we used 12 photons per time window — three times within a single burst are smoothed into only a single crossing of the threshold. The smoothed signal was analyzed for individual particle detections by locating the onset of each fluorescence burst, defined by the point at which the signal first exceeded the threshold.

We generated twenty 250ms fluorescence signals corresponding to a sample in the microfluidic geometry already described at an absolute concentration of 150pM. We approximate Ω as a 3.2μm cylinder with a radius of 350nm, in which case the relative concentration was 0.11. Again we approximated $j(t)$ for this geometry using its steady-state value, which we found to be $j(t) = 4.9\mu\text{s}^{-1}\text{pM}^{-1}$ by Monte-Carlo calculation. We used the extracted detection times to compute $\tilde{\Upsilon}_{corr}$, and the results are shown in part (a) of Fig. B.8. Along with the concentration estimates, we plotted error bounds corresponding to two standard deviations of the exact concentration estimator $\tilde{\Upsilon}$; we see, as expected, that at this concentration the detection-corrected estimator converges at a rate nearly equal to that of the idealized estimator $\tilde{\Upsilon}$.

In order to compare the performance of our concentration estimation technique with a standard method, we used FCS to analyze the same simulated signals that we analyzed with our detection-time method. In FCS, the sample concentration is proportional to the inverse of the autocorrelation curve $g(\tau)$ at $\tau = 0$ [46]. We calibrated our estimates by computing the FCS

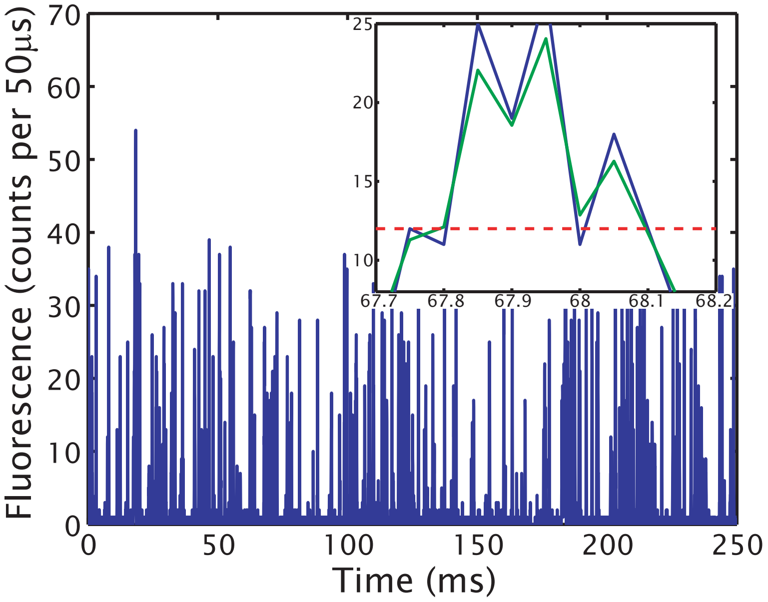


Figure B.7: Example fluorescence trace. The main plot shows a simulated 250ms fluorescence trace for a 150pM sample within a microfluidic channel. The inset compares the raw fluorescence signal (blue solid curve) with the output of the Lee filter (green solid curve) for a particular signal burst. The dashed red line indicates the fluorescence count threshold used in the analysis.

curves for seven 5s simulated fluorescence signals over a range of concentrations from 20pM to 400pM and finding the linear least squares fit to $g(0)^{-1}$. For a particular signal, we computed the FCS concentration estimate at any time T by generating the FCS curve of the signal for all photons detected at times $t < T$. In part (b) of Fig. B.8 we show the FCS concentration estimates computed in this way at 10ms intervals for each signal. The error bounds shown are identical to those in part (a) of the figure. These error bounds, though derived for $\tilde{\Upsilon}$, appear to be a good characterization of FCS as well.

B.5.3 Estimation of dynamic fluctuations in mean concentration

Next we demonstrate the ability of our technique to accurately estimate dynamically-fluctuating concentrations within the micro-channel geometry already described by simulating $10s$ fluorescence signals while varying the rate at which particles were injected at the inlet of the channel according to the sinusoid $C(t) = 200 + 100\sin(2\pi t/2.5)$ and the exponential $C(t) = 400\exp(-t/2.5)$ in order to modulate the concentration within the channel. Concentration estimates were made over 250ms time intervals. Error bounds of two standard deviations were computed for each time interval using Eq. B.22, and estimates were fit to sinusoidal and exponential functions. The results are shown in Fig. B.9.

The results in Fig. B.9 are quite promising. In both plots, nearly every estimate was within

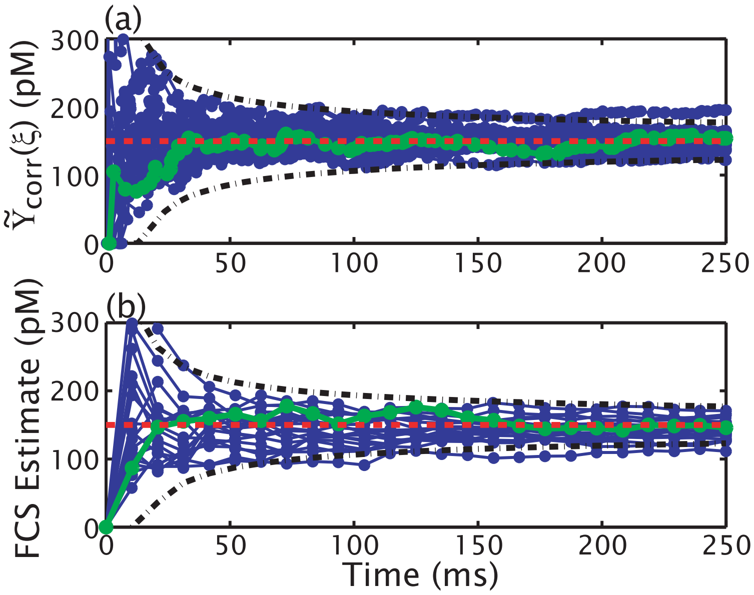


Figure B.8: Concentration estimates derived from fluorescence data. Twenty 250ms ensemble trajectories were simulated within a microchannel with a concentration of 150pM (red dashed line). Fluorescence signals were generated and used to compute concentration estimates (solid curves with markers) by (a) coupling the Lee filter with $\tilde{Y}_{corr}$, and (b) FCS. Markers indicate the times at which the estimates were updated. Green curves correspond to the fluorescence signal in Fig. B.7. 2σ error bounds (black dot-dash curves) were computed using Eq. B.25.

two standard deviations of the actual value of the concentration — in fact, the majority of the estimates were much closer than that. Furthermore, the numerical fits of the data to the functional forms of the concentration yielded curves that were almost identical to the actual concentration curves. While this success does depend on the fluctuations in $C(t)$ (the frequency of the sinusoid and the time constant of the exponential) being sufficiently slow, it shows that there are excellent prospects for measuring time-varying concentrations via SMD using our technique.

B.5.4 Simultaneous multiple-species concentration estimation

In our final set of simulations we illustrate how our detection-time concentration estimator may naturally be coupled with a single-molecule fluorescence burst analysis method in order to monitor the concentrations of several analyte species in a single solution. Until now, our approach in this paper has been limited to experiments in which only the total concentration of fluorescence-labeled species is measured, because no effort has been made to analyze the details of each fluorescence signal burst. A far more interesting approach is one in which there are several labeled species in solution that can be distinguished by their fluorescence fluctuation characteristics, and we examine each signal burst to determine which species was

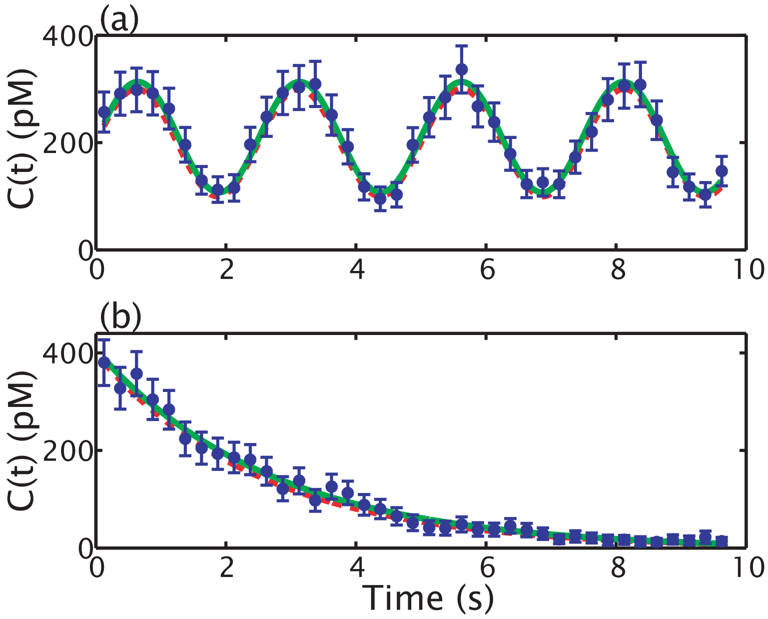


Figure B.9: Dynamic estimation of fluctuating concentrations. Diffusive trajectories were generated within a microchannel while modulating the sample concentration (red dashed curves) according to (a) $C(t) = 200 + 100\sin(2\pi t/2.5)$ and (b) $C(t) = 400\exp(-t/2.5)$. Fluorescence signals were generated and concentration estimates were made over 250ms time intervals by coupling the Lee filter with $\tilde{\Upsilon}_{corr}$. Blue markers indicate the concentration estimates over time. 2σ error bounds were computed according to Eq. B.22. The green curves are numerical fits of the data to the appropriate functional forms.

most likely to have caused it. We then compile lists of detection times for each species, and use our concentration estimator to determine the concentrations of those species over time.

In a previous paper we introduced a time-domain recursive Bayesian estimator for the analysis of individual single-molecule fluorescence bursts[16]. The estimator was based on an update rule for propagating probability densities corresponding to each species in solution over time. We showed that even in simple experiments — where the species in solution differed only in their diffusion coefficients — we could identify the species responsible for each burst quite accurately.

In order to illustrate the combined use of our identification and concentration estimation techniques, we simulated a mixture of freely-diffusing particles whose fluorescence characteristics are identical to those described in the previous two simulations, but whose diffusion coefficients are different, with $D_1 = 100\mu\text{m}^2\text{s}^{-1}$ and $D_2 = 33.3\mu\text{m}^2\text{s}^{-1}$. The detection region Ω was defined as the cylinder centered at the origin with radius 300nm and height 2μm, chosen because of the high intensity and spatial sensitivity of the excitation laser within this region. The concentration of the fast species was fixed at 20pM (0.0068 relative to Ω), and that of the slow species was fixed at 60pM (0.020 relative to Ω). Fluorescence signals were generated with sampling time 50μs, and the entire fluorescence signal was run through the recursive Bayesian estimator. The estimator propagated probability densities over a 3-dimensional cylindrical lattice with radius 1.25μm and height 6μm, with radial grid resolution of 50nm and axial resolution of 100nm. The detection times of individual particles were chosen to be the times at which the occupancy probability of Ω first exceeded 75%, and bursts were defined to have ended once that probability dropped below 10%. Recurrent detections were rejected by combining all bursts separated by less than 2ms, and the species corresponding to each set of recurrent bursts was chosen to be that with the greatest probability at the end of the final burst. Concentration estimates were made using the steady-state approximations $\tilde{y}_1 = 88\text{ns}^{-1}\text{pM}^{-1}$ and $\tilde{y}_2 = 180\text{ns}^{-1}\text{pM}^{-1}$, which were found by calibrating the concentration estimator on homogeneous samples of each species type. The results of the analysis of ten 10s fluorescence signals are shown in Fig. B.10.

The results in Fig. B.10 show an overestimate of the concentration of the slow species and an underestimate of that of the fast species. Such biases are unavoidable due to errors in identification, but their quantitative analysis will be left for future consideration. Even considering identification errors, though, these results are quite promising. Using time-domain techniques naturally suited to the problem we were able to extract information from these simulated signals that would have been difficult to obtain using any other approach. While simple, the binary sample that we simulated represents one of the most difficult experiments of this type in practice, as identification is much less sensitive to the diffusion coefficient than it is to differences

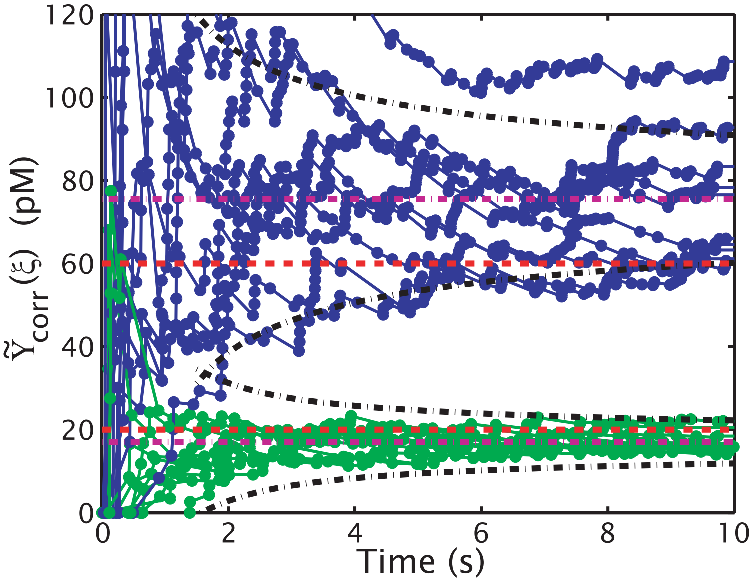


Figure B.10: Simultaneous concentration estimation of both species in a binary mixture. We simulated ten 10s ensembles of Brownian motion trajectories corresponding to a species with $D_1 = 33.3\mu m^2 s^{-1}$ at 60pM and $D_2 = 100\mu m^2 s^{-1}$ at 20pM. The resulting fluorescence signal was analyzed with a Bayesian estimator for species identification and $\tilde{Y}_{corr}$ for concentration estimates. The blue curve shows the estimates for species 1 and the green curve, for species 2. Markers indicate detection times. Red dashed lines indicate the sample concentrations. Violet dot-dashed lines indicate the mean value of the concentration estimates at the end of the simulation. 2σ error bounds (black dot-dash curves) about these mean concentrations were computed using Eq. B.25.

in the fluorescence spectrum, for example. We believe that the use of these time-domain methods in the future, possibly using more sensitive criteria for species identification, will make possible novel experiments in dilute chemical and biological kinetics.

B.6 Discussion

We presented an approach to concentration estimation in SMD experiments based upon the times between detections of individual analyte particles. In dealing only with detection times, we drastically reduced the SMD signal while retaining all information relevant to the sample concentration. We showed that the detection times have relatively simple statistics, which we explicitly computed using Smoluchowski's model for irreversible chemical kinetics. Using these statistics, we determined the minimal variance with which an arbitrary concentration estimation technique can converge over time — a calculation that would have been nearly impossible without using our reduced signal statistics. Furthermore, we derived a very simple time-domain concentration estimation technique, the convergence of which we explicitly showed was asymptotically optimal.

We presented numerical simulations to illustrate our technique applied in fluorescence-detection experiments in both free solution and in a microfluidic channel. We showed that the convergence of our technique is indistinguishable from that of FCS at low concentrations and thus, indirectly, we established quantitatively the rate at which FCS concentration estimates converge in time. We showed that our technique can be applied to accurately recover the functional form of slow concentration fluctuations from an SMD signal. Finally we showed that the combined application of our concentration estimator with a time-domain burst analysis method is an accurate and natural technique for monitoring the concentrations of multiple species in solution simultaneously.

Our work in this paper completes, in a sense, the work that we presented in our previous paper on Bayesian analysis of SMD signals. Using these techniques to monitor changing concentrations of multiple species in solution, we believe that novel experiments in chemical kinetics of dilute systems will be possible. One of the strengths of this time-domain approach is that by analyzing each signal burst individually we avoid the time-averaging that is a characteristic of other methods for SMD. This feature is critical for measuring the slow kinetics in dilute systems such as real and synthetic biological systems and microfluidic devices, all of which are of major interest to modern biotechnology, because in such systems it is unlikely that a reaction will occur over the course of just a single signal burst.

Another strength of our concentration estimation technique is that it provides a statistically rigorous measure of the uncertainty in its estimates in the form of an exact expression for the estimator variance. For this reason this technique is ideally suited to analytical experiments in which it is essential to provide accurate error bounds for measured concentrations.

In addition to kinetics experiments, our concentration estimation technique is ideal for efficiently measuring trace amounts of a single analyte, as might be required by low-power lab-on-a-chip devices. We see no reason to use a technique other than ours for this purpose, as its convergence is nearly optimal and it is computationally trivial to implement.

We must stress that essentially none of our presentation depends specifically on fluorescence detection as the SMD technique chosen. Our estimator only requires that the detection method have single-molecule resolution and that the detection rate scales linearly with concentration. We suspect that it will find wide application among the growing number of experimental methods for single-molecule detection.

Bibliography

[1] Douglas Magde, Elliot Elson, and W. W. Webb. Thermodynamic Fluctuations in a Reacting System — Measurement by Fluorescence Correlation Spectroscopy. *Phys. Rev. Lett.*, 29 (11):705–708, 1972. 1, 7, 49, 54, 161

[2] E. Brooks Shera, Newton K. Seitzinger, Lloyd M. Davis, Richard A. Keller, and Steven A. Soper. Detection of single fluorescent molecules. *Chem. Phys. Lett.*, 174:553–557, 1990. 1, 161

[3] Shimon Weiss. Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy. *Nat. Struct. Biol.*, 7(9):724–729, 2000. 1

[4] Ahmet Yildiz, Joseph N. Forkey, Sean A. McKinney, Taekjip Ha, Yale E. Goldman, and Paul R. Selvin. Myosin V walks hand-over-hand: single fluorophore imaging with 1.5nm localization. *Science*, 300:2061–2065, 2003. 1

[5] Ahmet Yildiz, Michio Tomishige, Ronald D. Vale, and Paul R. Selvin. Kinesin walks hand-over-hand. *Science*, 303:676–678, 2004. 1

[6] Daniel Nettels, Irina V. Gopich, Armin Hoffmann, and Benjamin Schuler. Ultrafast dynamics of protein collapse from single-molecule photon statistics. *Proc. Natl. Acad. Sci. U.S.A.*, 104:2655–2660, 2007. 1

[7] Jan Kubelka, James Hofrichter, and William A. Eaton. The protein folding 'speed limit'. *Curr. Opin. Struct. Biol.*, 14:76–88, 2004. 1

[8] N G. van Kampen. *Stochastic Processes in Physics and Chemistry*. Elsevier B.V., Amsterdam, revised and enlarged edition, 1992. 2, 15, 30, 31, 33, 42, 44, 45, 47, 49, 50, 51, 58, 103, 105, 123, 138, 187

[9] Jörg Enderlein. Tracking of fluorescent molecules diffusing within membranes. *Appl. Phys. B*, 71:773–777, 2000. 3, 5, 12

[10] Jörg Enderlein. Positional and temporal accuracy of single molecule tracking. *Single Mol.*, 1(3):225–230, 2000. 3, 5, 12

[11] Valeria Levi, QiaoQiao Ruan, K. Kis-Petikova, and Enrico Gratton. Scanning FCS, a novel method for three-dimensional particle tracking. *Biochem. Soc. Trans.*, 31:997–1000, 2003. 3, 12, 19

[12] Andrew J. Berglund and Hideo Mabuchi. Feedback controller design for tracking a single fluorescent molecule. *App. Phys. B.*, 78:653, 2004. 3, 5

[13] M. D. Armani, S. V. Chaudhary, R. Probst, and B. Shapiro. Using feedback control of microflows to independently steer multiple particles. *IEEE J. Microelectromech. Syst.*, 15: 945–956, 2006. 3, 10

[14] Adam E. Cohen and W. E. Moerner. Method for trapping and manipulating nanoscale objects in solution. *Appl. Phys. Lett.*, 88:223901, 2005. 3, 10

[15] Hu Cang, Chung M. Wong, C. Shan Xu, Abbas H. Rizvi, and Haw Yang. Confocal three dimensional tracking of a single nanoparticle with concurrent spectroscopic readouts. *Appl. Phys. Lett.*, 88:223901, 2006. 3, 10, 11

[16] Kevin McHale, Andrew J. Berglund, and Hideo Mabuchi. Bayesian estimation for species identification in single-molecule fluorescence microscopy. *Biophys J.*, 86:3409–3422, 2004. 4, 7, 185, 195, 205

[17] Andrew J. Berglund, Andrew C. Doherty, and Hideo Mabuchi. Photon Statistics and Dynamics of Fluorescence Resonance Energy Transfer. *Phys. Rev. Lett.*, 89:068101, 2002. 5, 54, 165, 168

[18] Andrew J. Berglund and Hideo Mabuchi. Tracking-FCS: Fluorescence correlation spectroscopy of individual particles. *Opt. Express*, 13:8069–8082, 2005. 5, 12, 49

[19] Andrew J. Berglund, Kevin McHale, and Hideo Mabuchi. Feedback localization of freely diffusing fluorescent particles near the optical shot-noise limit. *Opt. Lett.*, 32:145–147, 2007. 6, 89, 90, 97

[20] Kevin McHale, Andrew J. Berglund, and Hideo Mabuchi. Quantum dot photon statistics measured by three-dimensional particle tracking. *Nano Lett.*, 7:3535–3539, 2007. 6, 103

[21] D. Lumma, S. Keller, T. Vilgis, and J. O. Rädler. Dynamics of large semiflexible chains probed by fluorescence correlation spectroscopy. *Phys. Rev. Lett.*, 90:218301, 2003. 6, 54, 131, 143, 144, 148, 154

[22] Roman Shusterman, Sergey Alon, Tatyana Gavrinyov, and Oleg Krichevsky. Monomer dynamics in double- and single-stranded DNA polymers. *Phys. Rev. Lett.*, 92:048303, 2004. 6, 131, 144, 149, 150

[23] R. G. Winkler. Intramolecular dynamics of linear macromolecules by fluorescence correlation spectroscopy. *Phys. Rev. E*, 73:041919, 2006. 6, 131, 134, 141, 146, 154, 156

[24] E. P. Petrov, T. Ohrt, R. G. Winkler, and P. Schwille. Diffusion and segmental dynamics of double-stranded DNA. *Phys. Rev. Lett.*, 97:258101, 2006. 6, 131, 149, 150, 158, 159

[25] Adam E. Cohen and W. E. Moerner. Internal mechanical response of a polymer in solution. *Phys. Rev. Lett.*, 98:116001, 2007. 6, 131, 143, 144, 149, 155

[26] Elliot L. Elson and Douglas Magde. Fluorescence Correlation Spectroscopy. I. Conceptual Basis and Theory. *Biopolymers*, 13:1–27, 1974. 7, 161, 184

[27] Anthony E. Siegman. *Lasers*. University Science Books, Sausalito, CA, 1986. 9, 103

[28] S. Chaudhary and B. Shapiro. Arbitrary steering of multiple particles independently in an electro-osmotically driven microfluidic system. *IEEE Trans. Contr. Syst. Technol.*, 14: 669–680, 2006. 10

[29] Adam E. Cohen and W. E. Moerner. Control of nanoparticles with arbitrary two-dimensional force fields. *Phys. Rev. Lett.*, 86:093109, 2005. 10

[30] Adam E. Cohen and W. E. Moerner. Suppressing Brownian motion of individual biomolecules in solution. *Proc. Natl. Acad. Sci. U.S.A.*, 103:4362–4365, 2006. 10, 11

[31] H. Cang, C. S. Xu, D. Montiel, and H. Yang. Guiding a confocal microscope by single fluorescent particles. *Opt. Lett.*, 88:223901, 2007. 10, 11

[32] Timothy Ragan, Hayden Huang, Peter So, and Enrico Gratton. 3-D particle tracking on a two-photon microscope. *J. Fluoresc.*, 16:325–336, 2006. 11, 19

[33] William H. Press. Flicker noises in astronomy and elsewhere. *Comment. Astrophys.*, 7: 103–119, 1978. 11

[34] Per Bak, Chao Tang, and Kurt Wiesenfeld. Self-organized criticality: an explanation of $1/f$ noise. *Phys. Rev. Lett.*, 59:381–384, 1987. 11

[35] Valeria Levi, QiaoQiao Ruan, M. Plutz, S. Belmont, and Enrico Gratton. Chromatin dynamics in interphase cells revealed by tracking in a two-photon excitation microscope. *Biophys. J.*, 89:4275–4285, 2005. 12, 19

[36] Andrew J. Berglund and Hideo Mabuchi. Performance bounds on single-particle tracking by fluorescence modulation. *Appl. Phys. B*, 83:127–133, 2006. 12, 18, 42, 89

[37] Andrew J. Berglund. *Feedback Control of Brownian Motion for Single-Particle Fluorescence Spectroscopy*. PhD thesis, California Institute of Technology, 2006. 15, 18, 35, 37, 39, 42, 49, 57, 61, 76, 89, 92

[38] S. Prajna, A. Papachristodoulou, P. Seiler, and P. A. Parrilo. *SOSTOOLS: Sum of squares optimization toolbox for MATLAB*, 2004. 23

[39] S. Chandrasekhar. Stochastic Problems in Physics and Astronomy. *Rev. Mod. Phys.*, 15 (1):1–89, 1943. 30, 62

[40] Edward Nelson. *Dynamical Theories of Brownian Motion*. Princeton University Press, second edition, 2001. 30

[41] D. J. Tritton. *Physical Fluid Dynamics*. Oxford University Press, Oxford, 1988. 32

[42] Andrew J. Berglund, Kevin McHale, and Hideo Mabuchi. Fluctuations in closed-loop fluorescent particle tracking. *Opt. Express*, 15:7752–7773, 2007. 35, 37, 39, 49, 57, 61, 64, 68, 90, 111

[43] Bernard Friedland. *Control System Design: An Introduction to State-Space Methods*. McGraw-Hill, Inc., 1986. 35, 36

[44] John Doyle, Bruce Francis, and Allen Tannenbaum. *Feedback Control Theory*. Macmillan Publishing Co., 1990. 37, 38

[45] Sidney Redner. *A Guide to First-Passage Processes*. Cambridge University Press, Cambridge, UK, 2001. 44

[46] Oleg Krichevsky and Grégoire Bonnet. Fluorescence correlation spectroscopy: the technique and its applications. *Rep. Prog. Phys.*, 65:251–297, 2002. 51, 52, 161, 201

[47] Sören Doose, James M. Tsay, Fabien Pinaud, and Shimon Weiss. Comparison of photophysical and colloidal properties of biocompatible semiconductor nanocrystals using fluorescence correlation spectroscopy. *Anal. Chem.*, 77:2235–2242, 2005. 54, 102

[48] Dennis E. Koppel. Statistical accuracy in fluorescence correlation spectroscopy. *Phys. Rev. A*, 10:1938–1945, 1974. 56

[49] Michael E. Fisher. Shape of a self-avoiding walk or polymer chain. *J. Chem. Phys.*, 44: 616–622, 1966. 57

[50] J. Rička and Th. Binkert. Direct measurement of a distinct correlation function by fluorescence cross correlation. *Phys. Rev. A*, 39:2646–2652, 1989. 64

[51] Adam E. Cohen. *Trapping and Manipulating Single Molecules in Solution*. PhD thesis, Stanford University, 2006. 65, 89

[52] Robert W. Boyd. *Nonlinear Optics*. Elsevier Science, San Diego, California, second edition, 2003. 70

[53] R. Hanbury Brown and R. Q. Twiss. Correlation between photons in two coherent beams of light. *Nature*, 177:27–29, 1956. 75, 101

[54] R. Hanbury Brown and R. Q. Twiss. The question of correlation between photons in coherent light rays. *Nature*, 178:1447–1448, 1956. 75, 101

[55] P. Horowitz and W. Hill. *The Art of Electronics*. Cambridge University Press, Cambridge, UK, second edition, 1989. 85, 86, 91

[56] Pat Brown. *Poly-l-Lysine coating of slides*, 1999. 92

[57] R. Shankar. *Principles of Quantum Mechanics*. Springer, New York, 1994. 99

[58] A. P. Alivisatos. Semiconductor clusters, nanocrystals, and quantum dots. *Science*, 271: 933–937, 1996. 99, 100

[59] A. P. Alivisatos. Perspectives on the physical chemistry of semiconductor nanocrystals. *J. Phys. Chem.*, 100:13226–13239, 1996. 99, 100

[60] A. I. Ekimov and A. A. Onushchenko. Quantum size effect in the optical spectra of semiconductor microcrystals. *Sov. Phys.-Semicond.*, 16:775–778, 1982. 100

[61] A. L. Efros and Al. L. Efros. Interband absorption of light in a semiconductor sphere. *Sov. Phys.-Semicond.*, 16:772, 1982. 100

[62] L. E. Brus. A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *J. Chem. Phys.*, 79:5566–5571, 1983. 100

[63] C. B. Murray, D. J. Norris, and M. G. Bawendi. Synthesis and characterization of nearly monodisperse CdE (E=S, Se, Te) semiconductor nanocrystallites. *J. Am. Chem. Soc.*, 115: 8706–8715, 1994. 100

[64] Margaret A. Hines and Phillippe Guyot-Sionnest. Synthesis and characterization of strongly-luminescing ZnS-capped CdSe nanocrystals. *J. Phys. Chem.*, 100:468–471, 1996. 100

[65] P. Reiss, Joël Bleuse, and Adam Pron. Highly luminescent CdSe/ZnSe Core/Shell Nanocrystals of low size dispersion. *Nano Lett.*, 2:781–784, 2002. 100

[66] M. Jr. Bruchez, M. Moronne, P. Gin, S. Weiss, and A. P. Alivisatos. Semiconductor Nanocrystals as Fluorescent Biological Labels. *Science*, 281:2013–2016, September 1998. 100

[67] Warren C. W. Chan and Shuming Nie. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science*, 281:2016–2018, 1998. 100

[68] Daniel R. Larson, Warren R. Zipfel, Rebecca M. Williams, Stephen W. Clark, Marcel P. Bruchez, Frank W. Wise, and Watt W. Webb. Water-soluble quantum dots for multiphoton fluorescence imaging in vivo. *Science*, 300:1434–1436, 2003. 100

[69] X. Michalet, F. F. Pinaud, L. A. Bentolila, J. M. Tsay, S. Doose, J. J. Li, G. Sundaresan, A. M. Wu, S. S. Gambhir, and S. Weiss. Quantum Dots for Live Cells, in Vivo Imaging, and Diagnostics. *Science*, 307:538–544, January 2005. 100

[70] M. Dahan, T. Laurence, F. Pinaud, D. S. Chemla, A. P. Alivisatos, M. Sauer, and S. Weiss. Time-gated biological imaging by use of colloidal quantum dots. *Opt. Lett.*, 26:825–827, 2001. 100

[71] M. Nirmal, B. O. Dabbousi, M. G. Bawendi, J. J. Macklin, J. K. Trautman, T. D. Harris, and L. E. Brus. Fluorescence intermittency in single cadmium selenide nanocrystals. *Nature*, 383:802–804, 1996. 101, 122, 123

[72] Al. L Efros and M. Rosen. Random telegraph signal in the photoluminescence intensity of a single quantum dot. *Phys. Rev. Lett.*, 78:1110–1113, 1997. 101, 122, 123

[73] M. Kuno, D. P. Fromm, H. F. Hamann, A. Gallagher, and D. J. Nesbitt. Nonexponential "blinking" kinetics of single CdSe quantum dots: a universal power law behavior. *J. Chem. Phys.*, 112:3117–3120, 2000. 101, 124, 128

[74] K. T. Shimizu, R. G. Neuhauser, C. A. Leatherdale, S. A. Empedocles, W. K. Woo, and M. G. Bawendi. Blinking statistics in single semiconductor nanocrystal quantum dots. *Phys. Rev. B*, 63:205316, 2001. 101, 124, 128

[75] Jie Yao, Daniel R. Larson, Harshad D. Vishwasrao, Warren R. Zipfel, and Watt W. Webb. Blinking and nonradiant dark fraction of water-soluble quantum dots in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.*, 102:14284–14289, 2005. 101, 126

[76] Sungchul Hohng and Taekjip Ha. Near-complete suppression of quantum dot blinking in ambient conditions. *J. Am. Chem. Soc.*, 126:1324–1325, 2004. 101, 102, 110, 125, 127, 129

[77] B. Lounis, H. A. Bechtel, D. Gerion, P. Alivisatos, and W. E. Moerner. Photon antibunching in single CdSe/ZnS quantum dot fluorescence. *Chem. Phys. Lett.*, 329:399–404, 2000. 101, 102, 113, 114, 116

[78] Marlan O. Scully and M. Suhail Zubairy. *Quantum Optics*. Cambridge University Press, Cambridge, UK, 1997. 101

[79] H. J. Kimble, M. Dagenais, and L. Mandel. Photon antibunching in resonance fluorescence. *Phys. Rev. Lett.*, 39:691–695, 1977. 101

[80] Leonard Mandel and Emil Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, Cambridge, UK, 1995. 101, 109

[81] L. L. Hench and D. E. Clark. Physical chemistry of glass surfaces. *J. Non-Cryst. Solids*, 28: 83–105, 1978. 102, 114

[82] Crispin W. Gardiner. *Handbook of Stochastic Methods*. Springer-Verlag, Berlin Heidelberg, 2004. 105

[83] Arch W. Naylor and George R. Sell. *Linear Operator Theory in Engineering and Science*, volume 40 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2000. 109

[84] P. Michler, A. Imamoğlu, M. D. Mason, P. J. Carson, G. F. Strouse, and S. K. Buratto. Quantum correlation among photons from a single quantum dot at room temperature. *Nature*, 406:968–970, August 2000. 112, 113, 122

[85] George Casella and Roger L. Berger. *Statistical Inference*. Duxbury Press, Pacific Grove, CA, second edition, 2002. 115, 190, 191

[86] T. Takagahara. Biexciton states in semiconductor quantum dots and their nonlinear optical properties. *Phys. Rev. B*, 39:10206–10231, 1989. 119

[87] E. H. S. Burhop. The Auger Effect. *Proc. R. Soc. A*, 148:272–284, 1935. 122

[88] Andrew J. Berglund. Nonexponential statistics of fluorescence photobleaching. *J. Chem. Phys.*, 121:2899–2903, 2004. 124

[89] Matthew Pelton, Glenna Smith, Norbert F. Scherer, and Rudolph A. Marcus. Evidence for a diffusion-controlled mechanism for fluorescence blinking of colloidal quantum dots. *Proc. Natl. Acad. Sci. U.S.A.*, 104:14249–14254, 2007. 124, 128

[90] Rogier Verberk and Michel Orrit. Photon statistics in the fluorescence of single molecules and nanocrystals: correlation functions versus distributions of on- and off-times. *J. Chem. Phys.*, 119:2214–2222, 2003. 128

[91] M. Doi and S. F. Edwards. *The Theory of Polymer Dynamics*. International Series of Monographs on Physics. Clarendon Press, Oxford, 1986. 296. 131, 132, 133, 134, 135, 137, 139, 140, 141, 142, 143, 171

[92] Pierre-Gilles de Gennes. *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca, New York, 1979. 132

[93] Paul J. Flory. The configuration of real polymer chains. *J. Chem. Phys.*, 17:303–310, 1949. 134

[94] Melvin Lax. Classical noise IV: Langevin methods. *Rev. Mod. Phys.*, 38:541–566, 1966. 135

[95] Prince E. Rouse, Jr. A theory of the linear viscoelastic properties of dilute solutions of coiling polymers. *J. Chem. Phys.*, 21:1272–1280, 1953. 136

[96] Bruno H. Zimm. Dynamics of polymer molecules in dilute solution: viscoelasticity, flow birefringence and dielectric loss. *J. Chem. Phys.*, 24:269–278, 1956. 139, 140

[97] Richard P. Feynman, Robert B. Leighton, and Matthew Sands. *Feynman lectures on physics. Volume 2: Mainly electromagnetism and matter*. Addison-Wesley, Reading, MA, 1964. 140

[98] O. Kratky and G. Porod. Röntgenuntersuchung gelöster fadenmoleküle. *Rec. Trav. Chim. Pays-Bas*, 68:1106–1123, 1949. 141

[99] Roland G. Winkler, Peter Reineker, and Ludger Harnau. Models and equilibrium properties of stiff molecular chains. *J. Chem. Phys.*, 101:8119–8129, 1994. 141

[100] Paul J. Hagerman. Flexibility of DNA. *Ann. Rev. Biophys. Biophys. Chem.*, 17:265–286, 1988. 141

[101] P.-G. de Gennes. Quasi-elastic scattering of neutrons by dilute polymer solutions: I. free-draining limit. *Physics*, 3:37–45, 1967. 142

[102] E. Dubois-Violette and P.-G. de Gennes. Quasi-elastic scattering by dilute, ideal, polymer solutions: Ii. effects of hydrodynamic interactions. *Physics*, 3:181–198, 1967. 142, 144, 150

[103] Roland G. Winkler. Diffusion and segmental dynamics of rodlike molecules by fluorescence correlation spectroscopy. *J. Chem. Phys.*, 127:054904, 2007. 146

[104] Douglas E. Smith, Thomas T. Perkins, and Steven Chu. Dynamical scaling of DNA diffusion coefficients. *Macromolecules*, 1996:1372–1373, 1996. 154

[105] S. Weiss. Fluorescence Spectroscopy of Single Biomolecules. *Science*, 283:1676–1683, 1999. 160, 183

[106] Markus Rarbach, Ulrich Kettling, Andre Koltermann, and Manfred Eigen. Dual-Color Fluorescence Cross-Correlation Spectroscopy for Monitoring the Kinetics of Enzyme-Catalyzed Reactions. *Methods*, 24:104–116, 2001. 161

[107] Petra Schwille and Ulrich Kettling. Analyzing single protein molecules using optical methods. *Curr. Opin. Biotechnol.*, 12:382–386, 2001. 161

[108] Yan Chen, Joachim D. Müller, Peter T.C. So, and Enrico Gratton. The Photon Counting Histogram in Fluorescence Fluctuation Spectroscopy. *Biophys. J.*, 77:553–567, 1999. 161

[109] Peet Kask, Kaupo Palo, Dirk Ullmann, and Karsten Gall. Fluorescence-intensity distribution analysis and its application in biomolecular detection technology. *Proc. Natl. Acad. Sci. U.S.A.*, 96(24):13756–13761, 1999. 161

[110] Peet Kask, Kaupo Palo, Nicolas Fay, Leif Brand, Ülo Mets, Dirk Ullmann, Joern Jungmann, Johannes Pschorr, and Karsten Gall. Two-Dimensional Fluorescence Intensity Distribution Analysis: Theory and Applications. *Biophys. J.*, 78:1703–1713, 2000. 161

[111] Kaupo Palo, Ülo Mets, Stefan Jäger, Peet Kask, and Karsten Gall. Fluorescence Intensity Multiple Distributions Analysis: Concurrent Determination of Diffusion Times and Molecular Brightness. *Biophys. J.*, 79:2858–2866, 2000. 161, 176

[112] Ted A. Laurence, Achillefs N. Kapanidis, Xiangxu Kong, Daniel S. Chemla, and Shimon Weiss. Photon-Arrival-time Interval Distribution (PAID): a novel tool for analyzing molecular interactions. *J. Phys. Chem. B*, 108:3051–3067, 2004. 161, 171, 176

[113] Michael Prummer, Christian Hübner, Beate Sick, Bert Hecht, Alois Renn, and Urs P. Wild. Single-Molecule Identification by Spectrally and Time-Resolved Fluorescence Detection. *Anal. Chem.*, 72:443–447, 2000. 161

[114] Jörg Enderlein and Malte Köllner. Comparison between time-correlated single photon counting and fluorescence correlation spectroscopy in single molecule identification. *Bioimaging*, 6:3–13, 1998. 161, 175

[115] Jörg Enderlein and Markus Sauer. Optimal Algorithm for Single-Molecule Identification with Time-Correlated Single-Photon Counting. *J. Phys. Chem. A*, 105:48–53, 2001. 161

[116] Morris H Degroot. *Probability and Statistics*. Addison-Wesley Pub. Co., Reading, Massachusetts, 1986. 169, 173

[117] G. Bao and S. Suresh. Cell and molecular mechanics of biological materials. *Nature Materials*, 2:715–725, 2003. 169

[118] John Stockton, Michael Armen, and Hideo Mabuchi. Programmable logic devices in experimental quantum optics. *J. Opt. Soc. Am. B*, 19(12):3019–3027, 2002. 178

[119] M. Chaichian and A. Demichev. *Path Integrals in Physics Volume I: Stochastic Processes and Quantum Mechanics*. Series in Mathematical and Computational Physics. IOP Publishing LTD, London, 2001. 182

[120] Samuel T. Hess, Shaohui Huang, Ahmed Heikal, and Watt W. Webb. Biological and Chemical Applications of Fluorescence Correlation Spectroscopy: A Review. *Biochemistry*, 41 (3):697–705, 2002. 183

[121] Daniel Figeys and Devanand Pinto. Lab-on-a-Chip: A Revolution in Biological and Medical Sciences. *Anal. Chem.*, 72(9):330A–335A, 2000. 183

[122] Todd Thorsen, Sebastian J. Maerkl, and Stephen R. Quake. Microfluidic Large-Scale Integration. *Science*, 298:580–584, 2002. 183

[123] Andrew J. de Mello. Seeing single molecules. *Lab Chip*, 3:29N–34N, 2003. 183

[124] Petra S. Dittrich and Andreas Manz. Single-molecule fluorescence detection in microfluidic channels — the Holy Grail in μTAS? *Anal. Bioanal. Chem.*, 382:1771–1782, 2005. 183

[125] Hong Qian, Gary M. Raymond, and James B. Bassingthwaighte. Stochastic fractal behavior in concentration fluctuation and fluorescence correlation spectroscopy. *Biophys. Chem.*, 80:1–5, 1999. 184

[126] M. von Smoluchowski. Versuch einer mathematischen theorie der koagulationskinetik kolloider lösungen. *Z. Phys. Chem.*, 92:129–168, 1917. 187

[127] Attila Szabo. Theory of Diffusion-Influenced Fluorescence Quenching. *J. Phys. Chem.*, 93: 6929–6939, 1989. 187

[128] Raimund J. Ober, Sripad Ram, and E. Sally Ward. Localization Accuracy in Single-Molecule Microscopy. *Biophys. J.*, 86:1185–1200, 2004. 193

[129] Jörg Enderlein, David L. Robbins, Patrick W. Ambrose, and Richard A. Keller. Molecular shot noise, burst size distribution, and single-molecule detection in fluid flow: effects of multiple occupancy. *J. Phys. Chem. A*, 102(30):6089–6094, 1998. 195, 201

[130] Robert M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth. On the Lambert W Function. *Adv. Comput. Math.*, 5:329–359, 1996. 196

[131] I. I. Gikhman and A. V. Skorokhod. *Introduction to the Theory of Random Processes*. Dover Publications, Inc., Mineola, N.Y., 1996. 196

[132] M. Matsumoto and T. Nishimura. Mersenne Twister: A 623-dimensionally equidistributed uniform pseudorandom number generator. *ACM Trans. Model. Comput. Simul.*, 8:3–30, 1998. 198

[133] William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, UK, second edition, 1992. 198

[134] T. Geyer, C. Gorba, and V. Helms. Interfacing Brownian dynamics simulations. *J. Chem. Phys.*, 120(10):4573–4580, 2004. 198