DISSIPATIVE CONTROL AND IMAGING OF COLD ATOMS

by

JEREMY THORN

A DISSERTATION

Presented to the Department of Physics
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

June 2012

DISSERTATION APPROVAL PAGE

Student: Jeremy Thorn

Title: Dissipative Control and Imaging of Cold Atoms

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Physics by:

Jens Nöckel	Chair
Daniel Steck	Advisor
Michael Raymer	Member
Raghuveer Parthasarathy	Member
Michael Kellman	Outside Member

and

Kimberly Andrews Espy	Vice President for Research and Innovation/ Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded June 2012

DISSERTATION ABSTRACT

Jeremy Thorn
Doctor of Philosophy
Department of Physics
June 2012

Title: Dissipative Control and Imaging of Cold Atoms

We present an all-optical one-way barrier for cold rubidium 87 atoms. Along with the basic theory on how the barrier works, we describe the experimental setup and demonstrate an actual realization of the barrier. Such a barrier appears at first glance to violate the second law of thermodynamics; we examine that law and show explicitly that it is not violated for a general class of one-way barriers including our particular realization. As our lab is going to continue on to a different set of experiments requiring very sensitive imaging techniques, we finish with the development and application of a theory for comparing electron-multiplying charge-coupled device (EMCCD) cameras.

CURRICULUM VITAE

NAME OF AUTHOR: Jeremy Thorn

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, Oregon, USA
Whitman College, Walla Walla, Washington, USA

DEGREES AWARDED:

Doctor of Philosophy in Physics, 2012, University of Oregon
Bachelor of Arts in Mathematics, 2003, Whitman College
Bachelor of Arts in Physics, 2003, Whitman College

AREAS OF SPECIAL INTEREST:

Quantum Mechanics, Laboratory Control

PROFESSIONAL EXPERIENCE:

Research Assistant, Department of Physics, University of Oregon, Eugene, Oregon, USA, 2004–2012
Teaching Assistant and Tutor for Introductory Astronomy Course, Department of Physics, University of Oregon, Eugene, Oregon, USA, 2003–2004
Research Assistant, Department of Physics, Whitman College, Walla Walla, Washington, USA, 2002–2003
Calculus and Astronomy Tutor, Department of Physics, Whitman College, Walla Walla, Washington, USA, 2002–2003
Farm Machinery Operator, Dayton, Washington, USA, 1993–2011

GRANTS, AWARDS AND HONORS:

Henry V. Howe Scholarship Recipient, University of Oregon, Eugene, OR, 2008–2009

Hertz Foundation Finalist, Interview Stage, 2006

NSF Fellowship Honorable Mention, 2005

National Merit Scholar, 1999

PUBLICATIONS:

Elizabeth Schoene, Jeremy Thorn, Daniel Steck, *Cooling atoms with a moving one-way barrier*, Physical Review A **82** (2), 023419 (2010)

Peter Gaskell, Jeremy Thorn, Sequoia Alba, Daniel Steck, *An open-source, extensible system for laboratory timing and control*, Review of Scientific Instruments **80** (11), 115103 (2009)

Jeremy Thorn, Elizabeth Schoene, Tao Li, Daniel Steck, *Dynamics of cold atoms crossing a one-way barrier*, Physical Review A **79** (6), 063402 (2009)

Jeremy Thorn, Elizabeth Schoene, Tao Li, Daniel Steck, *Experimental Realization of an Optical One-Way Barrier for Neutral Atoms*, Physical Review Letters **100** (24), 240407 (2008)

Jeremy Thorn, Matthew Neel, Vinsunt Donato, Geoffrey Bergreen, Robert Davies, Mark Beck, *Observing the quantum behavior of light in an undergraduate laboratory*, American Journal of Physics **72** (9), 1210–1219 (2004)

TABLE OF CONTENTS

Chapter Page

LIST OF FIGURES

Figure Page

Figure Page

Figure Page

LIST OF TABLES

Table Page

CHAPTER I

INTRODUCTION

Physics builds on itself. In the classic tale of Sir Isaac Newton and the apple, Newton extended the idea that objects on Earth fall, and successfully described the motion of the planets. Planck used the established laws of thermodynamics and observations of blackbody spectra to first postulate that light was quantized. Physicists have now established that both light and matter are quantized and in certain limits may be thought of as interacting through fields such as Newtonian gravity. Building off of that, scientists have successfully described, probed, and discovered much about the world we live in, pushing the boundaries to what we know out to reveal yet more to discover.

One route taken by physics has resulted in precision measurements and control over the basic building blocks of matter: atoms [1–3]. While one could argue that we have always had access to atoms, isolating them for study required many advances in technologies. Atoms have very little mass and tend to move quickly; studying them or using them for measurements requires some way to hold them and, preferably, cool them down so that they may be observed over longer periods of time. The precision of lasers has been harnessed to cool atoms down to near absolute zero, and trap them for long periods of time. While initially worthy of a Nobel prize, these techniques have become common tools in cold atom laboratories [4–7]. A subsequent Nobel prize was awarded for the achievement of Bose-Einstein condensation in neutral atoms, a feat which used these same technologies as tools [8, 9]. Now, Bose-Einstein condensates themselves are rather common, and used as convenient vessels for further probes into interacting condensed matter systems and the weirdness of quantum mechanics [10–12].

Methods and techniques for trapping atoms are widely published. If carefully isolated, these atoms act according to the laws of quantum mechanics. They can be used to probe those laws, or to take advantage of them. For example, manybodied quantum-mechanical systems are often too small and evolve too quickly to be easily observed, and have too many degrees of freedom to be simulated on current computers. Systems of cold, isolated atoms follow similar rules and can have similar degrees of freedom, and there exist proposals for engineering cold atoms to simulate quantum systems [13]. The structure of atoms may be probed to reveal certain universal constants, and even check if perhaps these constants change over time [3]. Because there are so many atoms and molecules with different properties, it is inevitable that some experiments are best performed with certain atoms or molecules. For example, atoms with strong natural magnetic moments are useful for exploring interacting condensed matter systems [14]. It would be nice to be able to trap and cool any atom or molecule that would work well for a particular experiment. However, the relatively small list of atoms that have been cooled and trapped represents all that is easy to do with current cooling and trapping techniques, and adding to that list is difficult [14]. Currently, trapping molecules requires clever tricks, usually involving forming molecules from already-cooled atoms [15].

Our goals have been to further these pursuits in three ways. First, in cooperation with Mark Raizen's group at the University of Texas, Austin, we helped develop an alternative cooling method which could theoretically work on atoms and molecules for which current cooling techniques fail [16–21]. Our version uses an optical one-way barrier for cold atoms, which additionally functions as a realization of Maxwell's Demon [22]. Such a barrier may also be treated as the equivalent of an electronic diode for atoms, which may possibly be used in quantum computational logic gates [23–27].

Our second goal is to work towards probing open quantum systems to test and develop theories on how quantum measurements occur, and how they may be utilized to actively control quantum systems [28–31]. Aside from being interesting in itself, the ability to use feedback on a quantum system would aid in initial state preparation of experiments. It would also allow experimenters to correct for environmental perturbations, as opposed to passively damping them, enhancing our abilities to perform precise measurements [32, 33].

Finally, we hope to lower the barrier to entry in the study of cold atoms by helping develop flexible tools and techniques that are useful in many ways, particularly in cold-atoms laboratories. We emphasize publishing schematics and source code in an open-source manner so that others may continue development or modify our work to better suit their needs [34, 35].

In Chapter II of this thesis, we briefly describe the theory behind common methods for trapping and cooling atoms, providing derivations of simpler results and results where we can present the derivation differently than is typically done in literature and classes. Continuing through Chapter III, we describe our laboratory setup. Our discussion emphasizes the methods and hardware we use in our lab, some of the alignment tricks we used, and what types of equipment we use to achieve our goals. As previous theses from our lab cover many details of our setup, we omit some details and refer to the theses of Tao Li and Elizabeth Schoene in their absence [36, 37].

In Chapter IV, we describe our optical one-way barrier experiment, and demonstrate not only that it works to filter atoms in a uni-directional manner, but is also rather robust to perturbations in the precise setup. We outline how the relatively narrow splitting of the atomic states we used resulted in scattering being a more important effect than we had expected. This turned out to be both an impediment and a benefit. Our workaround, an asymmetric detuning, results in some excess heating in exchange for an added robustness to incorrect initial state preparation. We also show that some rather simplistic assumptions allow us to match the data from our experiments using computer simulations. Following that, we cover the thermodynamic implications of our one-way barrier, giving both a generic argument, and a specific computation that our one-way barrier, and any similar barrier mechanism, does not violate the second law of thermodynamics [18, 20]. While we did demonstrate that our one-way barrier, even with the extra scattering, could be used for cooling, we refer to Elizabeth Schoene's thesis for a discussion of that experiment [21, 37].

Having completed the one-way barrier experiment, we plan to use our setup to probe single cold atoms to test theories of quantum measurement and control.

Our plans require a method to image cold atoms with very little illumination, for which we plan to use electron-multiplying charge-coupled devices (EMCCDs). To that end, we also report our theory of noise in EMCCDs in Chapter V. In addition to providing a careful derivation of our theory from the general model of EMCCD operation [38–41], we demonstrate that it is surprisingly accurate and useful. We show that this theory can be used to determine the noise characteristics of a given camera, which we use to quantitatively compare various cameras. Our results may also be used in computer simulations, either to simulate the measurements from the camera, or in computing probabilities of events in quantum simultaions given the measurements from a real camera in an experiment. During this process, we discovered many potential pitfalls concerning EMCCD cameras, and we discuss a variety of things we learned to watch out for when checking cameras. These include such things as image artifacts, drifts in readout amplifiers, and various inaccuracies in the standard reported noise statistics.

CHAPTER II

BASIC TOOLS OF ATOM TRAPPING

In this chapter, we will describe the methods we use to trap and manipulate atoms. The methods described here have become quite standard in the field of cold atoms, and the formulas describing them can (and have) been derived in several ways. Here, we present the general methods with some simple methods of deriving formulas for them, along with some references for more in-depth treatments of the derivations. We will describe our particular setup in more detail in Chapter III.

Optical Molasses

There are several methods used for cooling atoms. Some of them, such as Zeeman slowers [4], cool mostly in one dimension, and others, such as supersonic expansion, produce a packet of atoms with a narrow spread in velocities, but a high center-of-mass velocity. One method in particular, known as *optical molasses* [5, 42], has become a popular mechanism for cooling cold, neutral atoms, possibly because it can easily be incorporated into a trap which achieves remarkably low temperatures (see Section II.2). The name refers to the fact that the light fields produce an effective drag force that opposes the velocity of the atom. Since this drag is a result of the Doppler shift, this is also known as *Doppler cooling.* The minimum temperature achievable by this method, which we will derive shortly, is known as the *Doppler limit*, or *Doppler temperature* [42–44].

The discovery and invention of optical molasses, as well as a few other cooling and trapping techniques such as the trap described in Section II.2 [5, 7, 43], led to the Nobel Prize in Physics being awarded to Steven Chu, Claude Cohen-Tannoudji, and William Phillips in 1997. Since then, optical molasses has become a common method for cooling and trapping atoms, and was instrumental in the achievement of Bose-Einstein condensation with neutral, weakly interacting atoms [8, 9], for which the Nobel Prize in Physics was awarded to Eric Cornell, Carl Wieman, and Wolfgang Ketterle in 2001.

The theory and implementation of optical molasses are both quite simple. Light carries momentum, and can therefore exert force on objects. The forces are small, and so tend to only affect small objects.[1]. We typically notice this force only in the way light can rattle atoms around in a solid, causing the solid to warm up. However, with something as simple as an isolated atom, that force can be controlled well enough to cool the atom, instead of heat it.

Optical molasses uses light that is closely tuned to a single atomic resonance. Near-resonant light reacts more strongly with the atoms than off-resonant light, just like shaking a ball on a spring produces a much larger reaction when the shaking is done at the resonant frequency as opposed to much higher or lower. This stronger reaction results in a larger force from the light. Working with near-resonant light also makes the effect simpler, as we can consider only the closest atomic resonance, all others being farther away and thus a weak effect.

We start with the scattering rate for a two-level atom illuminated with near-resonant light. As no two-level atoms actually exist [47], we interpret "near-resonance" as an atom being illuminated with a laser of angular frequency ω close enough to a particular atomic transition, with angular frequency ω_0, that the effects of all other transitions are negligible.

Using a standard two-level-atom dipole-coupling to an electromagnetic field, the equations of motion for the amplitudes in the two levels of the atom can be written out, and a decay rate can be phenomenologically added. With the decay term, the atom quickly settles down to a steady-state in the populations. The decay rate, Γ, is assumed to be the rate at which an excited atom decays down to the ground state. As such, we simply multiply it by the excited-state population in steady state

[1] There are discussions of using this force to propel spacecraft using a *solar sail*, which is a reflecting surface that is large and light enough that light from the sun would push on it hard enough to noticeably accelerate it [45]. More recently, similar effects are believed to account for certain accelerations experienced by the Pioneer 10 and 11 probes [46]

to achieve the steady-state scattering rate [48, 49]:

$$R = \frac{\Gamma}{2}\frac{\dfrac{I}{I_{\mathrm{sat}}}}{1 + \left(\dfrac{2\Delta}{\Gamma}\right)^2 + \dfrac{I}{I_{\mathrm{sat}}}}. \tag{II.1}$$

The detuning $\Delta := \omega - \omega_0$ is the difference in angular frequencies of the laser beam and the atomic resonance. It is positive if the laser is *blue-detuned* (at a higher frequency than the resonance frequency) and negative if the laser if *red-detuned.* The scattering rate, for low intensities, is proportional to the illuminating intensity I. However, as the intensity increases, the pumping rate of the atoms increases to the point where the atom is being raised to the excited state by the illuminating beam much faster than the state decays. However, the same illuminating beam also pumps the excited state back down to the ground state (called *stimulated emission*), and so the high-intensity limit is where the atoms is in the excited state half the time and the ground state the other half of the time. When multiplied by the decay rate, this results in a saturated (high-intensity) scattering-rate limit of $\Gamma/2$.

If we illuminate a cloud of atoms from all directions with red-detuned light, the atoms at rest will scatter light evenly from all beams, and so experience, on average, no net force. However, any beams the atoms move towards will be Doppler-shifted towards resonance, increasing the scattering rate from those beams. Those beams will then exert a larger force than the ones from which the atom is receding. As this happens regardless of the direction of motion, all the atoms experience the drag force commonly referred to as optical molasses. Since the drag is roughly proportional to the speed of the atoms, the resulting cooling rate reduces as the average kinetic energy decreases, until the cooling is balanced by the heating from the random scattering. In the high-intensity limit, where the scattering rate in Equation (II.1) saturates, the cooling rate is reduced by the broadening of the resonance, since the Doppler-shift in the detuning is masked by the intensity term in the denominator. As the heating rate is unaffected, the cooling limit is minimized when the light is below the saturation intensity, with an equilibrium kinetic energy given by [48, 49]:

$$K_{\mathrm{D}} = \frac{3\hbar\left(\Gamma^2 + 4\Delta^2\right)}{16\left|\Delta\right|}. \tag{II.2}$$

This is minimized when $\Delta = -\Gamma/2$, which results in the well-known Doppler limit:

$$k_{\mathrm{B}} T_{\mathrm{D}} = \frac{\hbar\Gamma}{2}, \tag{II.3}$$

where the relationship between the mean kinetic energy and the temperature is $K_{\mathrm{D}} :== \frac{3}{2} k_{\mathrm{B}} T_{\mathrm{D}}$.

We note that the Doppler limit is a lower bound only if we assume that Doppler cooling is the only cooling effect. As we will see in the next section, there are often other effects that can alter this limit. In fact, in the first attempt at three-dimensional optical molasses, the researchers were quite surprised at how far below the Doppler temperature their cooling methods reached [43]. They had to verify their findings using four different methods.

Optical molasses can be used to cool any two-level structure that couples to light. All that is needed is a symmetrical setup of laser beams from all directions, with the same intensity, and tuned slightly below the atomic resonance. If the beams are not all the same frequency, then if we can shift to a moving lab frame where they are, then the setup will act like optical molasses in that frame. Back in the lab frame, this results in cooling the atoms down to a state with a constant velocity (with some temperature relative to that). If the beams are not perfectly aligned, or the intensities are not perfectly matched, then the constant light forces do not perfectly cancel, and the result is a cooling force coupled with a constant acceleration, like gravity. While this can be used to counter the effects of gravity, it also means that slight mismatches in alignment and intensities are not catastrophic to the setup. Optical molasses without trapping is stable enough to achieve in the laboratory, with some care to reduce magnetic fields and get the beam intensities matched [5, 43].

The main problem with optical molasses is that atoms are not two-level systems. Hydrogen-like atoms, such as the rubidium atoms we use, have a hyperfine splitting in their ground state, from whether the spin of the nucleus aligns or anti-aligns with the electron angular momentum. Both of these states couple to the same excited states. This means that if the atom starts in one ground state, when excited, it might decay to the other state. Since optical molasses depends on many spontaneous emission events to work, the atom will end up in the wrong state quite quickly. Since the splitting of these two ground states tends to be far larger than the linewidth of the transitions, this puts the atoms too far out of resonance to be cooled further. The main solution to this is to use a *cycling transition*. This is a transition where the excited state has only a single decay path that leads back to the original ground state. In our case of rubidium 87, we use the transition from the $F = 2$ ground state to the $F' = 3$ excited state. Since the F quantum number can only change by 1 in a single-photon emission, a decay to the dark $F = 1$ ground state is dipole-forbidden. However, there is always the chance of an excitation to the nearby $F' = 2$ excited state, which is only a few hundred megahertz detuned. This latter state can decay to the $F = 1$ ground state, and, in fact, this happens often enough that a secondary solution to this problem is required.

The typical secondary solution to this is to have a beam resonant with a transition for the other ground state, to quickly pump the atom back to the state which is being cooled, although there are other solutions [14]. This second beam is referred to as the *repumping beam*. The necessity of dealing with other states in the atom is one of the main limitations of optical molasses. For atoms with a mostly three-level structure such as rubidium, Doppler cooling can be performed with just two laser beams. However, for more complicated structures, it is difficult to find cycling transitions. In addition, it is likely that any transition used for cooling will have many decay paths, each of which requires a beam to repump the atoms back to the cooling transition. Especially for molecules, which have many vibrational states, the number of laser frequencies (or other tricks) required quickly becomes infeasible. This is the reason why other cooling methods that do not require many spontaneous emission events, even if they perform worse than Doppler cooling, are useful.

Magneto-Optical Traps (MOTs)

Optical molasses is a great way to cool atoms, but, by itself, it cannot hold atoms in one place. However, with a clever configuration of magnetic fields and laser polarizations, we can add a trapping force to the effect. This combination is referred to as a magneto-optical trap, or MOT [7, 50].

Optical molasses provides a drag force that opposes the velocity of the atoms. In order to form a trap, we need to add a force that pushes towards some common location in space. The standard six-beam MOT setup is shown in Figure 2.1. It starts with three orthogonal pairs of counter-propagating beams that provide the optical molasses. If we pick each pair to have circular polarization, then they will couple to distinct magnetic sublevels of the atomic state. In particular, we pick both beams in a given pair to either be both right-handed, or both left-handed. Therefore, at a given location in space, their angular momentum will be pointed in opposite directions, and they will couple to different transitions as shown in Figure 2.2. When a magnetic field is applied, the excited states shift, changing the transition frequencies. For small magnetic fields, these shifts are linear in field strength.

As we mentioned at the end of Section II.1, if the cooling beams have different detunings, then we may be able to shift to a moving frame where the beams have the same detuning. For a single pair of counter-propagating beams, this is always the case. We simply move away from the beam that has a higher frequency (relative to the transition two which it is coupled). For small shifts, the detuning is linearly proportional to the speed, which we used to derive the Doppler cooling effect. In this moving frame, Doppler cooling happens in the normal manner, but back in the lab frame, the atoms are cooled to a state with a non-zero net velocity.

We will start with a magnetic field that is linearly dependent on the position and small enough that the shifts of the magnetic sublevels are linear with that field, and define the shifts as follows (we are still working in one dimension):

$$B = B_x x \tag{II.4}$$

$$\Delta E_e = g_e m_{F',e} B \tag{II.5}$$

$$\Delta E_g = g_g m_{F,g} B, \tag{II.6}$$

where we use g_g and g_e as the proportionality constants for the level shifts for a given magnetic field strength, and we use e (g) subscripts refer to the excited (ground)

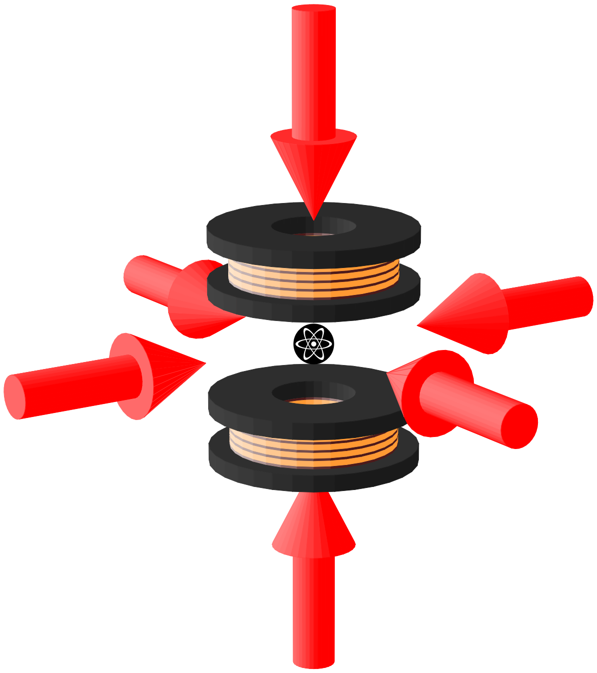


Figure 2.1. A simple schematic of a six-beam MOT setup. Three orthogonal pairs of counter-propagating beams provide the optical molasses while a pair of anti-Helmholtz coils provide a position-dependent Zeeman shift that causes the beams to also push the atoms to the center of the trap.

state of the cooling transition. The m_F factors refer to the magnetic-sublevel quantum number of the particular state we are using. The circularly polarized beams couple a given ground sublevel to an excited sublevel with a magnetic quantum number that is the sum of the ground-state quantum number with the angular momentum of a photon from the beam (± 1). Therefore, we can write $m_{F',e} = m_{F,g} \pm 1$, where we use the $+$ if the beam is circularly polarized along the $+x$ direction, and $-$ if the beam is polarized along the $-x$ direction. With this notation, the shift in the detuning of the beams is the difference in the shifts of the excited and ground sublevels:

$$\text{detuning shift} = \left[g_e \left(m_{F,g} \pm 1\right) - g_g m_{F,g}\right] B_x x. \tag{II.7}$$

While the overall detuning of the beams affects the final cooling temperature, we are looking for a force to push atoms towards $x = 0$, which requires the two counter-

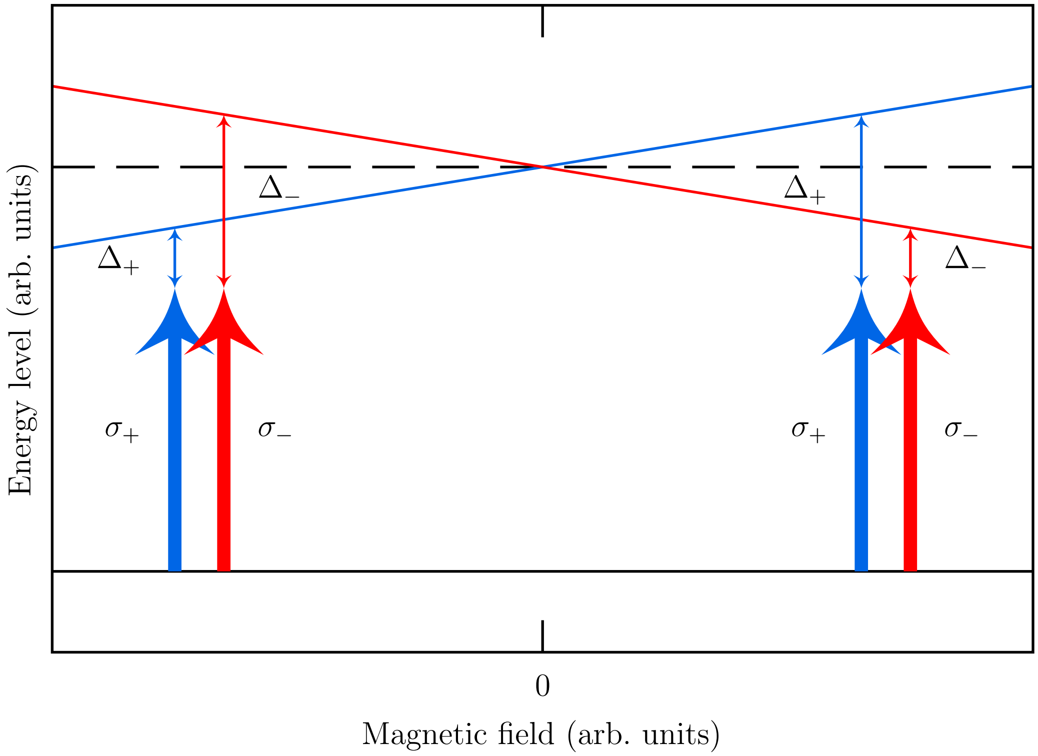


Figure 2.2. A simple schematic of the energy levels and beam polarizations used in a MOT. The excited-state Zeeman shifts are shown relative to the ground-state shifts for clarity. If counter-propagating beams have circular polarization with opposite projections onto the magnetic field (σ_+ and σ_-), then the states they couple to shift in opposite directions, effectively allowing the two beams to have different detunings (Δ_+ and Δ_-).

propagating beams to be detuned differently. Thus, we only care about the *relative* detuning shifts between the two beams:

$$\text{relative detuning shift} = 2g_e B_x x. \tag{II.8}$$

Both beams are below the unshifted atomic resonance, and so if $g_e B_x$ is positive, the level shifts will result in the beam with the $+x$ circular polarization being closer to resonance than the $-x$ beam, if $x > 0$. This means the $+x$ beam will push harder than the $-x$ beam. If both beams are left-handed (so the $+x$ beam is traveling in the $-x$ direction), then this will push the atoms towards the origin. Likewise, if $x < 0$, the $-x$ beam pushes harder, and the same configuration still pushes atoms towards the origin. If g_eB_x is negative, then using right-handed beams will always push towards the origin.

One way to think of Equation (II.8) is that a magnetic field shifts the rest frame of the trap. Assuming the beams each have the same detuning in the lab frame, then a magnetic field B along the beams shifts the states to which the beams are coupled so that the difference in the beam detuning is $2g_eB$. As the Doppler shift is $\overrightarrow{k} \cdot \overrightarrow{v}$ for angular frequencies, the beams will appear to have the same frequency in a frame moving with speed $2\pi g_eB/k$ relative to the laboratory frame, which will shift each beam by the same amount to cancel the relative difference. For the rubidium 87 D_2 line, g_e is about 0.93 MHz/G, and the wavenumber is about $k/2\pi = 1.26 \times 10^4$ cm^{-1} [51]. This means that the rest frame for optical molasses with a constant magnetic field is moving at a speed of about 74 cm/s per Gauss of magnetic field.

If we now expand to three dimensions, the situation becomes much more complicated because anywhere there is a magnetic field, it cannot be aligned with all six beams at once. Our simple argument where the sublevels of the atom in the magnetic field were also the eigenstates coupled by the circularly polarized light breaks down because the axis of circular polarization is no longer aligned (or anti-aligned) for all the beams. However, more complicated analyses show similar results to a simple three-dimensional generalization of our one-dimensional argument, and experiments show that a three-dimensional MOT works [50]. All beams are circularly polarized, each pair is either both left-handed or both right-handed, and we need only get a magnetic field configuration where the projection of the magnetic field along each of the three orthogonal beam axes is zero at the origin and is linearly related to the distance along that axis. This can be achieved with a pair of anti-Helmholtz coils. As shown in Figure 2.1, these are a pair of coils that share an axis, with the same current flowing between them, but in opposite directions. The result is that the magnetic fields cancel at the midpoint between the two coils.

The only thing we need to be careful about with the setup in Figure 2.1 is matching the handedness of the beam pairs with the direction of the magnetic fields. By the symmetry of the setup, the magnetic field has equal magnitude and direction radially from the origin. However, since the divergence of the magnetic field must be zero in the absence of magnetic monopoles, we have $B_x + B_y + B_z = 0$. With symmetry, $B_x = B_y$, and so $B_z = -B_x/2 = -B_y/2$. The magnitude is not as important as the sign, which is opposite along the axis of the anti-Helmholtz coils. This is why the radial beams are right-circularly polarized, while the axial beams are left-circularly polarized (or vice versa). Which is picked depends on the sign of g_e for the excited state, and which direction the current is flowing in the magnetic field coils.

Just as optical molasses is robust against various misalignments and intensity variations, MOTs are also robust against misalignments. They are actually more robust, since a small misalignment in optical molasses can launch atoms in a particular direction quite quickly, while, for a MOT, the restoring position force just causes the atoms to collect somewhere slightly shifted from the origin. By aligning Helmholtz or near-Hemlholtz coils around our traps, we can shift the region of zero field, effectively moving the trapping point around.

There is a variant of the six-beam MOT that is useful for its simplicity. Rather than have six separate beams, we can shine one large beam into a pyramid with a right-angle at the apex (a cone may be used as well) [52, 53]. Coupled with anti-Helmholtz coils aligned with the axis of the pyramid, this forms everything needed for a MOT. Parts of the beams reflect off of the sides of the pyramid and form the radial pairs of counter-propagating beams. When those beams reflect off of the sides again, the result is a beam that propagates against the incoming beam. This twice-reflected beam and the incoming beam form the axial beams. Assuming a small amount of phase shift with each reflection, a single reflection flips the circular-polarization of the beam, so that if the incoming beam is left-circularly polarized, all the radial beams are right-circularly polarized, and the outgoing beam is left-circularly polarized again, which is exactly what is needed for a MOT. Since high-reflection mirrors are typically dielectric stacks which do impart polarization-dependent phase shifts, the circular polarization tends to be slightly off. Even though we designed the coating on our mirrors to properly handle circular polarization at 45° incidence, they degraded some during the baking of our vacuum chamber. We compensated for this by adjusting the polarization of the input beam to slightly non-circular, although we found that it was not particularly sensitive to this.

The downside of a pyramid MOT is that it is hard to get good reflections at the edges and apex of the pyramid. The edges are not too much of a problem, as all the radial beams intersect in the center of the pyramid. The apex can be an issue, though, because that produces a hole in the outgoing axial beam. While this can be dealt with by shifting the center of the magnetic fields to be slightly off-axis to avoid that region, this setup is commonly used to produce a beam of cold atoms [54]. By simply leaving a hole at the apex, we create a MOT where atoms are cooled and collect in the center of the pyramid, where there is a beam missing. This results in a launching force that accelerates the atoms down the axis of the pyramid, and through the hole. As most of the atoms are relatively cool by the time they reach that axis, the result is a beam of atoms with very little spread in transverse position, transverse velocity, and longitudinal velocity.

We use a double-MOT setup, which has both a standard six-beam MOT and a pyramid MOT [55]. We have two competing requirements for our atom-trapping setup. The first is that we want to be able to trap and hold atoms undisturbed for as long as possible. The second is that we want to be able to quickly collect many atoms from the background. The first means that we want as few atoms floating around in the vacuum chamber as possible; the second means we want many. The double-MOT is a common solution to this. The vacuum chamber is split into two parts, each held at relatively good vacuum, and connected with a tube that is long and narrow enough that flow between the chambers is negligible. On one side of the chamber, we have a rubidium source that keeps that side at a relatively high background pressure of rubidium atoms. We have a pyramid MOT there, with a hole in the apex. This produces a beam of cold atoms that is launched through the narrow tube to the other side of the chamber, where the atoms are trapped in a standard six-beam MOT. As there is no rubidium source on that side of the chamber, and a substantial amount of vacuum-maintaining pumping power, the cold-atom beam is the main source of atoms on that side of the chamber. By turning the first MOT beams on, we produce an atomic beam that can load our second MOT very quickly; by turning the first MOT beams off, the untrapped atoms quickly disperse, and our trapped atoms can be dealt with in a region with very few background atoms.

We mentioned that MOTs tend to outperform the Doppler limit, and we will now briefly discuss why. The general effect is typically described in a one-dimensional case, as the full three-dimensional case, just as with describing how a MOT works, is complicated enough that the proof is often left to experiment. There are several variants of this cooling technique, generally referred to as *polarization-gradient cooling* [56]. One with counter-propagating linearly polarized beams, perpendicular to each other, is typically called *Sisyphus cooling*, and is the one more often described. The one that better describes what happens in a MOT is where the beams have the circular polarization used in the MOT. Either way, there is a position-dependent polarization. In the case of the MOT, the two circularly polarized beams interfere to produce a linear polarization that rotates around the beam axis when traveling along the beam axis. As an atom with multiple magnetic sublevels travels through such a gradient, the steady-state populations of the magnetic sublevels shift as the polarization changes. This effect tends to spin the polarization of the atom with the electric field polarization, and the rotating atom couples more strongly to one circular polarization than the other. It happens that, with a negative detuning, the beam opposing the atomic motion scatters more light than the other beam, and the result is an enhanced cooling effect [56]. The limit of this cooling effect has been observed to be on the order of the single-photon cooling limit, and well below the Doppler cooling limit. A calculation of the temperature limit for this cooling method shows that the minimum temperature is proportional to beam intensity, and decreases as the detuning *increases* [56].

In order to trap as many atoms from the background as possible, we actually want rather intense beams fairly close to resonance (about half a linewidth away in angular frequency, according to the optimal Doppler cooling detuning). After loading, we usually have a second stage where we reduce the intensity and frequency of the trapping lasers to cool the atoms using polarization gradient cooling. For the rubidium 87 in our lab the decay rate of the MOT transition is about 38 MHz, which corresponds to a Doppler cooling limit of about 300 μK. Once we trap the atoms, by reducing the frequency and intensity of the trapping beams in the MOT, we are able to achieve a temperature as low as about 30 μK. This is well below the Doppler limit, well above the recoil limit of 300 nK, and on the order of what the one-dimensional theory predicts [56].

Optical Dipole Traps

The MOT is a great way to initially trap and cool atoms, but the constant scattering of light by the atoms destroys coherences of the electronic states. They also constitute position measurements that collapse the position wavefunction of the atoms and introduce random momentum kicks. Since we intend to perform experiments that need atoms that are not perturbed by constant scattering of light, we need one more type of trap in our experiment.

The *optical dipole trap*, often called a *far-off-resonant trap* (FORT), exploits the conservative-potential effect of a far-off-resonant beam acting on a two-level atom [6, 42, 57]. An atom in an electric field is polarized by that field. If the field is oscillating, the dipole moment of the atom oscillates with the same frequency, with an amplitude and phase determined by the frequency (relative to the atomic resonance) and amplitude of the electric field oscillations. When the oscillations are far enough off resonance that scattering of light is negligible, this behaves much like a classical polarizable particle with a linear Hooke's-Law-like restoring force. Just like a mass on a spring, if the electric field oscillates faster than the resonant frequency, the dipole oscillates out of phase. If the electric field oscillates slower than the resonant frequency, the dipole moment oscillates with the electric field. It works out that when the atomic dipole oscillates with (against) the electric field, the atom is attracted to (repelled from) the higher intensity regions of the light field. The final result is that the atom appears to be affected by a conservative potential proportional to the beam intensity and inversely proportional to the detuning [48, 49]:

$$V \propto \frac{I}{\Delta}. \tag{II.9}$$

The detuning Δ is the difference between the laser angular frequency and the atomic resonance angular frequency, as used in Equation (II.1). In fact, for large detunings, we note that the scattering rate from Equation (II.1) has a similar formula,

$$R \propto \frac{I}{\Delta^2}, \tag{II.10}$$

except that the power of the detuning in the denominator is larger. This larger power means that the magnitude of the dipole potential can be held constant by increasing the intensity and detuning together, while simultaneously decreasing the scattering rate to negligible values. In this limit, the optical dipole effect acts like a nearly perfect conservative potential in which the atoms can move. Because the detunings involved are often quite large, it is often the case that the detuning is comparable across multiple atomic transitions, and even comparable to the optical frequencies themselves. In these limits, we must abandon the two-level atom approximation *and* the rotating wave approximation. The result is that a beam that is far-detuned from *all* resonances provides a conservative potential for atoms of the form [48, 49]:

$$V = \sum_i \frac{\hbar \Gamma_i^2 I}{8 I_{\mathrm{sat},i}} \left[\frac{1}{\omega - \omega_{0,i}} - \frac{1}{\omega + \omega_{0,i}} \right], \tag{II.11}$$

where the sum is over all the transitions of the atom from whatever ground state it happens to be in, and Γ_i, $\omega_{0,i}$, and $I_{\mathrm{sat},i}$ are the decay rate, angular frequency, and saturation intensity for a given transition, respectively.

Far-off-resonance light also has a non-mechanical effect on atoms, known as the *ac Stark shift*, the *light shift*, or the *Lamp shift* [58]. The potential in Equation (II.11) comes about because the eigenvalues of the Hamiltonian shift with position. However, those same shifts may be thought of as changing the state of the atoms. As a quick example, take a simplified version of the Hamiltonian in a light field:

$$\hat{H} = \hbar \begin{bmatrix} \omega_e & \Omega \\ \Omega^* & \omega_g \end{bmatrix}$$

This represents the Hamiltonian of an atom with excited-state and ground-state angular frequencies of ω_e and ω_g, respectively, coupling to an optical field with Rabi frequency Ω, the angular frequency at which the optical field cycles the atom from the ground to the excited state and back. The presence of the laser effectively mixes the excited and ground states. If we diagonalize the matrix, we find the eigenvalues are shifted from the original values of ω_e and ω_g. If we were to then illuminate the atom with another laser of a different frequency, and treat that as a perturbation, we could then analyze its effect in this diagonalized state. As with the one-beam case, the strength of the coupling is determined by the difference of the eigenvalues compared to the laser frequency, assuming the laser frequency is different enough from ther first laser frequency so that any beat frequencies are too high to have an effect. However, that difference between the eigenvalues is now changed by the first laser. The presence of one laser shining on an atom, when the effect is strong enough to create a dipole trap, shifts the resonant frequencies of the atom. Precisely what that shift is, both direction and magnitude, depends on the coupling of both the ground state and the excited state with all other states of the atom, but most generally, the effect is that beams that would be near resonance for an unperturbed atom are typically off resonance for a perturbed atom. The magnitude of the shift is also dependent on the intensity of the illuminating laser, and so for an atom trapped in a dipole trap, the exact shift of resonance depends on the location of the atom in the trapping beam.

The dipole trap we used in our experiment typically had a ground-state shift of -19 MHz and the focus. We calculated that the D_2 resonance was shifted by anywhere from 50 MHz to 100 MHz to the blue, depending on which excited state we used for the transition. While this is large compared to the transition linewidth, making it difficult to be exactly on resonance. For the beams we used in our one-way barrier, one was detuned by much more than these shifts, making the shifts a negligible effect. While the other beam was intended to be resonant with a certain transition, the shift was small enough that even though the beam was not on resonance, because it was so tightly focused it was intense enough to nearly saturate the atoms, and hence still quickly pump them to the other transition. The only time this shift was problematic for us was when we wanted to image the atoms. There, we did not have the intensity to compensate for this shift (and a variable shift would make a non-uniform conversion from intensity to atomic density). However, since imaging destroys the cloud of trapped atoms, it needs to be done at the end of an experiment, and so we could simply turn the trapping beam off for imaging.

Imaging

We now discuss the two methods of imaging cold atoms that we used in our lab. For discussing both of these imaging techniques, we note that our experiments typically have atomic densities small enough that the illuminating beams are not significantly attenuated as they pass through the atomic cloud. Under that assumption, all atoms are illuminated equally. Experiments that deal with high atomic densities, such as some BEC experiments, need to be more careful with their imaging techniques [59].

The first method is called *fluorescence imaging*, and basically involves turning on optical molasses. As the atoms scatter the light forming the optical molasses, they are simultaneously cooled, which keeps them from dispersing too rapidly, but also scattered that light in all directions. We simply focus a camera on the atoms and image them via the light they scatter. Imaging, or at least measuring, atoms by the light they scatter from optical molasses was used in the first optical molasses experiments, the first MOTs, and may be used in some BEC experiments [5, 7, 59]. Atoms in a MOT are particularly easy to image, as they scatter light the entire time they are trapped, while remaining confined to the trap.

Atoms that are not trapped in a MOT may also be imaged with fluorescence imaging. While the optical molasses effect does cool free atoms to keep them from dispersing too quickly, they do disperse, limiting the amount of time that the atoms can be illuminated. It is also important to cancel any magnetic fields when imaging free atoms. As discussed in Section II.2, magnetic fields in optical molasses make the “rest” frame for the molasses one that is moving in the lab frame. Using the 74 cm/s per Gauss derived in Section II.2 for rubidium 87, the Earth’s magnetic field, which is about 0.5 G, will result in the atoms moving on the order of 40 cm/s. Our beams are on the order of a centimeter, so without cancelling the Earth's magnetic field, we could only image the atoms for about 25 ms before they were thrown away, neglecting the time it takes for the atoms to accelerate to that speed. To prevent smearing beyond the expansion of atoms in optical molasses, we need to cancel magnetic fields to well below the one-Gauss level.

Just as with optical molasses, the optimal parameters for fluorescence imaging seem to consist of a laser frequency on the order of an atomic linewidth below the atomic resonance, with the intensity near, but not quite at, the saturation limit. The saturation limit is where the intensity term (squared Rabi frequency) in the denominator of the scattering rate in Equation (II.1) is comparable to the other terms in the denominator. This is also where the Doppler cooling limit was adversely affected. Essentially, the Rabi frequency should be smaller than the atomic linewidth, to prevent excess heating of the atoms. If the illuminating light is too dim, however, then there will not be a strong enough signal to image the atoms before they disperse. Using light near the saturation limit tends to be a decent starting point. For our setup, we found we could get a reasonably strong image using the same intensity and detuning that seems to form the best MOT, with a few tens of milliseconds of illumination (typical images were 10 to 20 ms). This seemed to be a good compromise between signal strength and dispersal of the atoms.

We used fluorescence imaging for number and temperature measurements in addition to direct imaging. The atoms are illuminated from all directions, with circularly polarized light, and so should emit approximately isotropically. Therefore, with a good estimate of the local beam power, the scattering rate given by Equation (II.1) gives the amount of light scattered each second by each atom (this gets easier if the light is above the saturation intensity, where the detuning matters less). A quick calculation can reveal what fraction of the emitted light the camera receives. If the camera is calibrated, then we can change that into an intensity measurement, and figure out the number of atoms at each point in the image.

The lens we used for imaging atoms for the majority of our experiments is a 55 mm lens with an aperture setting of $f/3.5$. This aperture is often the limiting case, not the actual diameter of the entrance to the lens. The diameter of the effective lens size, taking the aperture into account, is the focal length divided by the aperture setting, 55 mm/3.5 ≈ 15.7 mm. The camera is pointed nearly straight at the atoms, and so the aperture should appear to be a circular disk of radius R and a distance d from the atoms. The solid angle ratio (dividing by the 4π for full-coverage) subtended by that disk is:

$$\frac{\Omega}{4\pi} = \frac{1}{2}\left(1 - \frac{d}{\sqrt{d^2 + R^2}}\right). \tag{II.12}$$

Our aperture is approximately 8.75 in from the MOT (using the inch-spaced holes in our table), which is approximately 222 mm. Using Equation (II.12) with $d =$ 222 mm and $R = 15.7$ mm$/2 = 7.85$ mm, we find that we should be capturing about 0.03% of the emitted light. According to the data sheet for the sensor in our camera, we find that the quantum efficiency of the detector at 780 nm is about 50% [60]. We have a piece of RG715 Schott glass in front of the camera sensor to act as a filter, which reduces the efficiency to about 40%, making our total collection efficiency about 0.0125%. The atoms in our setup are nearly saturated, and so emit approximately $\left(38.1 \times 10^6\right)/2$ photons per second each, as the saturated limit in Equation (II.1) is the half the atomic lifetime. When multiplied by our collection efficiency, we find we should record between 2000 and 3000 photons per second per atom. That needs only be multiplied by the illumination time of the image (typically on the order of 20 ms) to get the number of photons the camera should receive per atom (in our example, about 40 to 60).

Knowing the conversion from camera values to actual power (or photon count) completes the scaling from camera values to number of atoms in each pixel of the final image. We calibrated our camera by shining light of known power directly into the camera. We were careful to make sure we would not approach the damage threshold of the camera sensor. The data sheet for our camera quotes a maximum number of electrons per pixel (1×10^5 e$^-$), at which point the sensor saturates [60]. That is a safe limit, and should be well below the damage threshold. Using the quantum efficiency of about 50% for our wavelength, we aimed to use an intensity that would saturate the image in about half a second. The pixels are about 9 μm on a side. Taken all together, that works out to a maximum intensity on the order of 1×10^{15} photons per second per square meter (picking a smaller number, just to be safe), which is about 0.26 mW/m^2, or 0.026 μW/cm^2. An intensity of this level should definitely be safe for the camera, especially with shorter exposure times. Our imaging system was set up for approximately $2:1$ imaging, so the spot size on the sensor should be the same order of magnitude as the size of a collimated beam entering the camera, and we can simply use a beam with a peak intensity of about 0.026 μW/cm^2. Once we get images with that, we can vary it to approach the actual saturation, and then take a series of images with various pulse lengths. We can compute the camera signal level by either fitting a Gaussian profile to the image, or by checking the values of individual pixels. Plotting these levels as a function of pulse length gives us a linear plot. By fitting a line to this plot and measuring the slope, we reduce the effect of background noise and shot-to-shot noise, and arrive at a value of about 6.40 photons per camera signal unit. This measurement was done with the Schott glass in place, and so this accounts for the transmissivity of the filters as well as the quantum efficiency. Thus, we can simply take numbers directly from the camera, multiply by 6.40 to get photons, and then divide by the number of photons we expect to see per atom to get number of atoms. As a word of warning, we have used different sets of filters with the camera. If things are changing often, this calibration should perhaps be done every now and then, just to check.

We also used fluorescence imaging to measure temperatures of trapped atoms. This is a rather simple measurement in which atoms are released from the MOT (by turning the lasers and magnetic fields off), allowed to fall for a known period of time, and then imaged. The rate at which the atom cloud disperses gives a measure of the temperature. If the atoms start as a point-source, then after a period of time t, the atoms will have a position distribution equal to t times the velocity distribution. The variance in one direction, $t^2 \langle v^2 \rangle$, yields the temperature through the relation $m \langle v^2 \rangle /2 = k_{\mathrm{B}} T/2$, where m is the mass of the atoms. While the atoms do not start as a point source, the expansion typically washes out the initial distribution quickly, once $\sqrt{t^2 \langle v^2 \rangle}$ is larger than the initial spread. This is because the final spatial distribution is a convolution of the initial distribution with the velocity distribution (multiplied by the time t), under the typically decent assumption that the velocity distribution of atoms is not dependent on the initial location. As shown in Section B.1, the variance of the final distribution is the sum of the variances of the two distributions being convolved. The variance would then work out to be $V^2 + t^2 \langle v^2 \rangle$, where V^2 is the initial spatial variance. We can arrive at $\langle v^2 \rangle$ and compute the temperature by using a quadratic fit of the variance in t. We can also arrive at $\langle v^2 \rangle$ by computing the standard deviation, waiting until it becomes linear in t (when the velocity distribution has washed out the initial spatial dependence), and using the squared slope of that. This technique is known as the *time-of-flight* method, and is a common temperature measurement in the cold atom field [5, 59].

When doing a temperature measurement, it is important that the atoms do not spread out too much during the imaging itself. For a given exposure time, the amount of spreading during the exposure is constant, and so the convolution adds a fixed variance to the distribution, which is equivalent to having a larger initial spread. That means that as long as the imaging does not distort things too terribly (by accelerating the atoms, for instance) and is a constant effect from image to image, it will not affect the temperature measurement.

Sometimes we wish to image with high resolution, such as when observing atoms trapped in our dipole trap. The size of small clouds of atoms trapped in the dipole trap are such that the diffusion during a normal fluorescence image substantially alters the shape of the cloud. While this diffusion can be reduced by decreasing the imaging time, the signal also decreases with imaging time and there is a limit at which no more information can be obtained.

Our second type of imaging gets around this difficulty. This method is called *absorption imaging*, and rather than imaging the light emitted by the atoms, we measure its absence. Absorption imaging is one of the preferred methods of imaging BECs due to their typically high optical density [59], but is also useful with the lower densities with which we work. As shown by our rough calculation on the fraction of emitted light that is collected by the camera, the vast majority of emitted light in fluorescence imaging is not imaged. While we cannot alter the emission pattern much in free space, we can get around this by instead trying to lose all the light. In absorption imaging, the atoms are illuminated by a single beam, preferably with a flat intensity profile, that shines directly into the camera. The atoms scatter light in random directions, and cast a shadow in that beam. Previously, we used Equation (II.12) to compute that our camera collects about 0.03% of the light scattered by the atoms. Since the camera collects so little of the scattered light, we can assume that all the light scattered by the atoms is lost to the camera. As with fluorescence imaging, we typically want to be near, but below, the saturation intensity of the atoms. This is because below saturation, the amount of light the atoms scatter is proportional to the amount of light there is. Therefore, regardless of the incoming intensity, the atoms scatter the same fraction of light. This makes absorption imaging somewhat immune to having an intensity profile that is not constnat (something that can be a problem with fluorescence imaging). Once the beams pass the saturation intensity, the atoms stop emitting more light, and so the amount of light scattered becomes a constant (still proportional to the number of atoms) in an increasing background, making the measurement noisier.

Since we are sending a laser beam directly into the camera, it is important that the beam is weak enough so as not to damage the camera. We used the same threshold discussed earlier for calibrating the camera as a starting point, but note that it is quite conservative. After checking our pulse-length control, we started with that intensity and increased the intensity by a few orders of magnitude, from about 0.02 μW/cm^2 to about 2 mW/cm^2 (note the change from microwatts to milliwatts), which is roughly the saturation intensity for rubidium 87. When we acquire an image, we want to use the full dynamic range of the camera, in order to get the strongest signal-to-noise ratio, where signal is the amount of light received and noise is mostly from the electronics within the camera, and largely independent of the signal received. We also want to be able to image the atoms as quickly as possible, and so a larger intensity is best. Operating near (but under) the saturation intensity for rubidium 87, saturating the camera took on the order of 10 μs. With such a fast imaging time, atoms had very little time to move, and so, for us, absorption imaging allowed us to see atoms without smearing them much.

We actually tried to be well below the saturation limit, but the same order of magnitude. For this discussion, we will assume we are well below the saturation limit, in which case the scattering rate in Equation (II.1) or Equation (II.10) is proportional to the beam intensity. By multiplying by the energy of a photon, we can convert a scattering rate to a power loss, giving the fraction of power not captured by the camera. If the camera collects a sufficient fraction ϵ of the scattered light, we can correct for that by also multiplying by $1-\epsilon$. We will call that fraction of power not captured L, for loss. Assuming that all the atoms are in a small enough volume that any power scattered by one decreases the amount of power available for the next, the ratio of power remaining after passing N atoms is $r = (1-L)^N$. We can measure that power ratio, and, knowning L, find the number of atoms as $N = \ln(r)/\ln(1-L)$.

In absorption imaging, we take two images, one with atoms, and one without, and compare them. We look at the fraction of light missing in the one with the atoms, and divide by the amount of light present, and that ratio gives us the number of atoms. For rubidium, the maximum scattering rate for an atom is one the order of 10 MHz. A single photon has an energy on the order of an electron-Volt, and so this corresponds to a power loss on the order of 10^{-12} W per atom. A single pixel in our images represents a square on the order of 20 μm on a side in the focal plane, which corresponds to about 4×10^{-6} cm^2, so it would take on the order of 10^3 atoms in the space of one pixel on our camera to scatter a large fraction of the light for that pixel. We typically only have a few thousand atoms trapped when we are performing absorption imaging, and they are spread over many pixels. Therefore, we expected, and our images confirmed, that a tiny fraction of the illuminating light is scattered when we perform absorption imaging. With so little power lost, we can use $r = 1-\epsilon$ with $\epsilon \ll 1$. In that case, the number of atoms reduces to $N \approx -\epsilon/\ln(1-L) \approx \epsilon/L$, where the last approximation takes advantage of the fact that L is small, too. Either way, the result is that the fractional power loss, in the low-density limit, is linearly proportional to the number of atoms, and if we want to maximize the signal for a given number of atoms (make ϵ larger), we need to make sure the loss rate is as large as possible. As quick look at the scattering rate formula tells us that we need to have the beams be resonant ($\Delta = 0$) for that, and so for absorption imaging, we try our best to make sure the beams are resonant. In practice, we checked this by acquiring images for a range of detunings, and finding the detuning that produced the strongest signal.

While the fraction of light scattered in absorption imaging is independent of the intensity of the light, computing how much light is present requires that the total fluence of light (time-integrated intensity) is the same between the images. The acousto-optic modulator we used to shutter the beam tended to produce temperature-related fluctuations in the intensity of our illumination pulses. To combat that, we had to add a separate circuit that would monitor part of the beam intensity, which we describe in Section III.8.

To summarize the pros and cons of the two types of imaging we use, fluorescence imaging allows for a stronger signal, but at the expense of the atoms spreading out during the image. This makes fluorescence imaging better for imaging smaller numbers of atoms (and for imaging diffuse spreads, such as in temperature measurements). Fluorescence imaging is also rather sensitive to differences in intensity. Absorption imaging allows for exposures that are fast enough to avoid spreading of the atoms, but tends to have a worse signal-to-noise ratio. This makes it more useful where spatial resolution is important.

In both of these types of imaging, we have the repumping beam on. The repumping beams are so much weaker than the main imaging beams that their contributions to the scattered light should be negligible. In addition to the power discrepancy, the repumping beams only scatter light when an atom decays to the $F = 1$ ground state. Because the beams we used are tuned to a transition that should not be able to decay to that state, this is a relatively rare occurrence, and suppresses the effect of the repumping beam even more.

When imaging, we needed to have our trapping beam off. As described earlier, the ac Stark shift for atoms in our dipole trap is significant, and shifted the atoms well out of resonance. This prevents atoms within the beam from being imaging, but the fix is simple. Since imaging the atoms kicks them enough that the experiment cannot continue past the imaging, we simply turned the dipole trap beam off, usually about 0.5 ms before turning the imaging beams on. That was easily long enough for the beam to be fully off, without letting the atoms spread too much. Oftentimes, we left the main barrier beams used in our one-way barrier experiment on during imaging. This was because the mechanical shutter we used to block that beam did not have sub-millisecond repeatability. Leaving the beam unblocked prevented this from being something that changed from shot-to-shot. The ac Stark shift from this beam resulted in a very narrow black streak in the image where the beam crossed the trapped atoms. Rather than being a problem, this allowed us to find where the barrier beams crossed through the trap, and we used this in alignment.

We finally note that these imaging techniques can be crudely modified to get an idea of how many atoms are in a particular ground state. Because both imaging techniques use a single transition, any atoms that are not in the state with that resonance do not scatter light, and so are invisible. We combat that by having a repumping beam on during the imaging (and perhaps shortly beforehand). By simply leaving that beam off, the short absorption beam pulse will only react with atoms in one of the states ($F = 2$, for the beams we use). However, we need to be careful that the pulse does not pump atoms quickly to the dark state ($F = 1$ for our setup). During the short pulse length, we can reasonably expect most of the atoms in the $F = 2$ state to remain in that state, but the fact that the atoms do change limits the accuracy of this measurement. We can calibrate this by taking absorption images with the repumping beam on shortly beforehand (to put all the atoms into the $F = 2$ state), and comparing that to an image where we leave the repumping beam on throughout. The differences in the two images tell us what fraction of the atoms get pumped to $F = 1$ during such an image. We can then leave the repumping beam off altogether to get an idea of what fraction of atoms are in the $F = 2$ state. While we only used this method with absorption imaging, it should also work with fluorescence imaging. However, over the many millisecond pulse required for fluorescence imaging, we cannot expect the atoms to stay in the $F = 2$ state. This will limit the effective pulse length, making for a weaker signal.

CHAPTER III

EXPERIMENTAL HARDWARE

We now devote some time to describing the particular setup we used for our experiments. As most of the hardware is rather standard in cold atom laboratories, and other theses from our lab delve into the details, construction, and part numbers of our setup [36, 37], we simply try to give an overall picture of the setup.

The idea of the setup is very simple: We want some lasers of the right frequency to shine on the atoms in our vacuum chamber. The devil is in the details, though, and there are many details, making for a very complicated setup. Each laser, while cheap, requires a cavity to help set the frequency, and some other hardware to set the right physical properties. We also need some absolute frequency reference, some way to amplify laser power, magnetic field coils, a separate dipole-trap laser, an ultra-high vacuum chamber, and some way to precisely control everything. We cover these items in this chapter.

Vacuum Chamber

Neutral atom traps can typically only hold atoms with sub-kelvin temperatures, often millikelvins and below; a collision with a background atom at room temperature would knock any trapped atom out for good [61]. In order to achieve long lifetimes for trapped atoms, it is necessary to have an excellent vacuum [7]. In this chapter, we outline our methods for building an atom-trapping-grade vacuum, and refer to other theses for more information on the specific parts used for our chamber [36, 37].

Our baking procedure was based off of procedures and ideas from multiple sources [62–64]. The main components of the chamber are standard vacuum parts, mostly machined from 304L or 316L stainless steel. The custom-built parts were machined from 316 stainless steel, using Trim-sol as a lubricant, because it is free of silicone and sulfur. Vacuum parts fresh from the factory were considered clean enough to jump straight to the baking; all other parts with surfaces that would be exposed to the vacuum were rinsed or sonicated with, in order: an alconox solution, deionized water, spectroscopic-grade acetone, and spectroscopic-grade methanol. We finish with methanol because it is less likely to leave residues than acetone. After the cleaning, the parts were then dried with dry nitrogen boil-off from a liquid nitrogen dewar so as not to introduce any water or contaminants to the metal. The parts were then baked at 480°C for 48 hours. This bakes trapped hydrogen out of the metal, and forms an oxide coating that helps seal the remaining hydrogen inside the bulk of the metal. All connections were designed to have vents, so that any trapped air pockets could be pumped out more easily. When handling parts prior to assembly, we wore powder-free nitrile gloves which we changed often. Non-metal pieces, such as windows, were typically not baked at this stage. Our ion pumps were factory-baked before being sealed, and so we did not bake them again.

We assembled the chamber with one small ion pump on the higher pressure side (where the pyramid MOT would load from a higher rubidium 87 background). On the lower-pressure side, where the six-beam MOT would be, we have a much larger ion pump for general pumping and pressure measurements, a titanium sublimation pump, and five titanium-powder getter pumps. All of the pumps are located so that there is no line-of-sight path from the pumps to the trapping regions of the chamber; this is so that when we heat the titanium sublimation pump or the getters, they will not spray the trapping area with metallic vapor or other contaminants. We also have a fused-silica cell made by Hellma attached to the chamber; this is the actual science area where the six-beam MOT is. This Hellma cell is a long piece with square cross-section, made of four optically-contacted pieces of fused silica. The pieces are 5 mm thick, making the Hellma cell 30 mm across on the outside, and 20 mm across on the inside. One end is sealed off with a fused silica end plate, while the other opens into the rest of the vacuum chamber. Having a fully transparent cell allows for a lot of optical access, which is important to us. Inside the chamber, we also mounted a pair of mirrors that can deflect a beam coming in from the Hellma cell out two other windows. This is so that the dipole trap beam will not heat the sides of the vacuum chamber, which would cause out-gassing and raise the pressure.

Once assembled, we hooked up a turbo pump to the chamber and performed a helium leak-check. The turbo pump connected to the chamber in two places, on either side of the differential pumping tube that divides the higher-pressure side from the lower-pressure side. Once we were satisfied that we had no leaks, we wrapped the chamber with heater tape and built an oven around the chamber. The oven was made from ceramic fire bricks wrapped in clean, oil-free aluminum foil (to keep the dust from the bricks down), with a few clean aluminum and steel sheets to distribute weight on the bricks and provide structure for the top of the oven. We then baked the chamber at around 200°C for over a week, while continuously pumping on it with the turbo pump. The bake temperature was limited by the maximum temperature several of the anti-reflection coatings could handle, and the cables attached to the ion pumps.[1] The pressure was monitored continuously via a gauge on the turbo pump station, and the temperature was monitored via several thermocouples distributed through various places in the oven. We ran the titanium sublimation pump and the getter pumps throughout the bake to keep them warm, occasionally flashing them to higher temperatures. Initially, we kept the ion pumps off, until peak temperature was reached and the pressure seemed to have leveled off, at which point we turned them on.

After baking, we performed another helium leak check, tightening any connection that seemed to have even the smallest leak. We then closed the bake-able vacuum valves where the turbo pump station attached to the chamber, and disconnected the turbo pump.

Once the chamber is sealed and baked, we cannot break vacuum without needing to re-bake the chamber. That meant that the rubidium we planned to use needed to be inside the chamber prior to baking. We had a glass ampoule containing approximately 2 g of rubidium inside it, which was sealed inside a tube attached to the chamber through a bake-able valve. The valve was open during the bake, but

[1] We actually baked the chamber several times. Once was a test, as we had not actually added the Hellma cell. We also had a few leaks become apparent during baking, and once a power outage caused the turbo pump station to vent air into the chamber, requiring more baking afterwards.

allows us to close off the rubidium source if the pressure in the chamber ever gets too high. After we were satisfied the chamber was finished, we broke the glass and released the rubidium. For this purpose, we had a metal rod inside the chamber attached to a flexible bellows. This formed a battering ram we could use to smash the rubidium ampoule without breaking vacuum.

Our chamber design has the center of the pyramid MOT 7.25" above the surface of the table. That atomic beam travels horizontally through the differential pumping tube, which is vertically offset from the higher vacuum side of the chamber. The centerline of the six-beam MOT, which corresponds to the axis of the Hellma cell, is 7" above the surface of the table. We originally assumed the atomic beam would accelerate until the atoms were shifted out of resonance of the pushing beam, limiting the atoms to a speed corresponding to a Doppler shift of a few linewidths. If we assume the atoms would be accelerated approximately 3 linewidths out of resonance, then the beams would drop approximately 1/4" due to gravity between the two MOTs. We can do a quick check of this to see our mistake. As discussed in Section II.1, a laser beam exerts a force on an atom proportional to the scattering rate. If we work in the limit where the atom is moving fast enough that the detuning is the dominant term in Equation (II.1), then, since the detuning is proportional to the speed v, the acceleration of the atom is roughly proportional to $1/v^2$. Writing this as $\mathrm{d}v/\mathrm{d}t \propto v^{-2}$, we can take advantage of the chain rule to move differentials around to arrive at $v^2\,\mathrm{d}v \propto \mathrm{d}t$.[2] Integrating yields $v^3 \propto t$, to within a constant, or $v \propto t^{1/3}$. The important point here is that, because of the long polynomial tails in the scattering rate, the velocity does *not* saturate after a few linewidths, but keeps accelerating. When we solving the differential equations via the same methods, but using the full scattering-rate equation, we see that the atoms would accelerate to a detuning of 3 linewidths on the order of a few milliseconds, during which time they would hardly have traveled a few centimeters, much less than our approximation. It is likely that the beam travels much faster than we predicted, and so sinks very

[2] Without using differentials, that would be $v^2\,\mathrm{d}v/\mathrm{d}t$ is equal to a constant. By the chain rule, the time-derivative of $v^3/3$ is the derivative with respect to v, multiplied by $\mathrm{d}v/\mathrm{d}t$, which is exactly what we have. We then integrate both sides with respect to time, and get $v^3/3$ on one side, and t on the other (with a constant thrown in somewhere).

little due to gravity.

We have found that our mistake is to our benefit. Having the six-beam MOT below the atomic beam is useful because the atoms trapped in that MOT are not hit by either the atomic beam or the light that accelerates that beam. By raising the six-beam MOT temporarily during loading, using the magnetic fields, we can get decent trapping rates and still be easily outside the main beam. Besides, if the beam accelerated atoms over the entire length of the pumping tube, the beams of the six-beam MOT would not have a long enough interaction time to slow the atoms back down to trappable speeds. Catching the edges of the beam that are likely accelerated less is probably to our benefit.

External Cavity Diode Lasers (ECDLs)

One of the reasons we use rubidium 87 is that certain diodes used in CDROM drives are very near the resonance of the rubidium 87 D_2 line, making diode lasers for rubidium 87 cheap and easy to find. Semiconductor diode lasers are generally relatively cheap and typically do not need maintenance or adjustments. The gain medium, being a solid material, has a fairly broad emission range, which allows them to be tuned over a relatively large spectrum (several nanometers, for our diodes). That, coupled with the fact that they are small and so have a very short cavity, means they tend to drift in frequency and can be very sensitive to back-reflections, temperature, and other environmental factors. They are also quite susceptible to static electricity and current surges.

In order to trap atoms, we need to be able to set the frequency of the lasers to within about a linewidth (6 MHz for rubidium 87) of the atomic transition, and have it be stable to well below that limit. This requires being able to tune the laser to a particular frequency and lock it there, something that cannot be done with a free-running laser diode.

To combat these problems, we built an enclosure with an extended cavity around the diodes. This setup is typical for labs working with laser diodes, and has earned the acronym ECDL, for external cavity diode laser [65–67]. The “cavity” for a stand-alone diode is typically the back of the diode, which reflects, and the front of the diode, which transmits. Since the front of the diode is a weak reflector, as it usually only reflects because the index of refraction does not match that of the surrounding air, we can put a reflective surface in front of the diode to make an extended cavity which overrides the front surface. While the front of the diode can be anti-reflection coated to make an ECDL more stable, this is not required, and we have not needed to do this.

We use a Littrow configuration for our lasers [67]. In this design, a diffraction grating is used as a replacement for the front of the cavity. The grating is blazed (meaning the individual surfaces are tilted) to enhance the first-order diffraction back into the diode laser, while suppressing other diffractions. A straight reflection is used as the exit port. To provide a tunable cavity, both the length of the cavity and the angle of the grating must select the same wavelength. Assuming the grating is set so that these conditions are satisfied at one point, it turns out that mounting the grating on an arm where the pivot is in the plane of the diode (with the beam as the normal) satisfies both of these conditions over a wide range of motion. This allows us to tune the cavity with a piezo that bends the arm by small amounts.

The entire setup is enclosed to help reduce acoustic noise and to allow thermal stabilization. Tuning the temperature and current through the laser helps select the peak gain frequency of the laser. The actual diode lasers we have tend to lase at 784 nm without a cavity. Through a combination of the laser cavity and keeping the lasers at a temperature below room temperature (which tends to shorten their emitting wavelength), we are able to get them to lase in resonance with the 780 nm D_2 line of rubidium 87. If a laser is set up well, we can sweep it over the 6.8 GHz range of the rubidium 87 D_2 hyperfine structure without mode hops. Since we rarely need the full spectrum, and it usually takes a lot longer to find the right combination of temperature, operating current, and grating alignment to get that quality of behavior, our lasers usually cannot sweep through the entire spectrum reliably without mode hops, but they are at least on that order.

The lasers are relatively stable against vibrations, but careful inspection of the error signal when locked to the transitions of rubidium 87 show that they do pick up voices of people talking in the lab. They can typically stay locked to such a transition for on the order of a day, but can be knocked off lock by dropping a wrench on the optical table. A separate group in our lab has developed a superior laser design that is more stable, much less susceptible to vibrations, and has a much narrower inherent linewidth [68].

Finally, every laser on our table requires some extra optics, shown in Figure 3.1. The diodes output a linearly polarized beam, and we rotate the diodes so the polarization is vertical, which helps prevent changing the polarization when the beam reflects off of the grating. That beam exits the enclosure through a Brewster-angle window to reduce reflective loss, and then reflects off of two steering mirrors. These mirrors allow us to get a coarse re-alignment of the laser should we ever need to replace it.[3] The diode lasers emit an elliptical beam, which we pass through an anamorphic prism pair which squeezes the beam into a more circular shape. We then pass the beam through a half-wave plate to rotate the beam to 45° polarization, so it can pass through an optical isolator (which rotates the beam by 45° to return it to vertical polarization). The isolator insulates the laser against back-reflections further down the optical path. It is also used as an injection port on the slave lasers, described later in Section III.4. To differentiate these lasers from the slave lasers used to amplify light, we refer to the ECDLs as the *master lasers*.

Saturation Absorption Spectroscopy (SAS)

In order to form a MOT, we need the lasers to be stable to less than a linewidth (about 6 MHz). Since the base frequency corresponding to the wavelength of 780 nm is about 380 THz, this requires a frequency stability of better than one part in 10^7, preferably 10^8. Since we cannot depend on hitting that blindly, and the lasers are not stable to that degree of precision, we actively lock our lasers to some rubidium samples. However, room-temperature Doppler shifts are on the order of 1 GHz, which is much larger than we can tolerate. Fortunately, there is a clever trick to

[3] Various irises throughout our optical table aid in this realignment, although if such a realignment is ever necessary, we still need to fine-tune alignments throughout the full optical path.

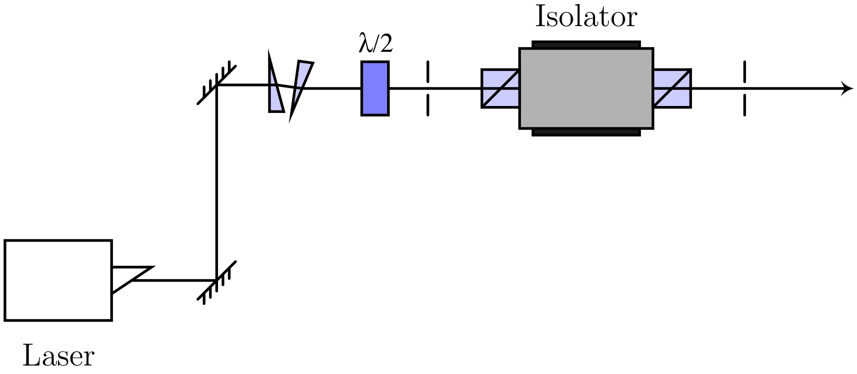


Figure 3.1. A schematic of the optics that are present on every one of our lasers. The beam exits through a Brewster-angle window, reflects off of two steering mirrors, passes through an anamorphic prism pair, a half-wave plate, and an optical isolator. We have irises placed before and after the isolator to help with beam alignment.

cancel the Doppler shifts.

The clever trick is called *saturated absorption spectroscopy* (SAS) [66, 69–72]. While there are several ways to cancel out the thermal motion of the atoms [73–75], we will describe the version we use in our lab. The general idea is we pass two beams through a sample of rubidium, with the two beams traveling in opposite directions. In the rest frame of any given atom, those two beams may or may not be of the same frequency. If the beams are both resonant with a transition of that atom, the atom will scatter light from both beams. Ordinarily, both beams would then be attenuated by an amount proportional to the scattering rate, which is proportional to the beam intensity. However, as can be seen in Equation (II.1), the atom can saturate. If one beam is much stronger than the saturation intensity, it pumps the atom hard enough that half of the time the atom will be in the excited state at any given time. This means that the second beam will experience less scattering than if the stronger beam were not there.

The setup, then, has three beams. Two of the beams, called *probe beams*, are below the saturation intensity, and pass through equal amounts of rubidium vapor. If they are of the same frequency and intensity, the rubidium will scatter the same amount from each beam. Subtracting the two beams will then give a constant signal as a function of frequency. A third beam, called the *pump beam*, is above (or near) the saturation intensity, and crosses one (but not both) of the probe beams, in the opposite direction. The pump beam should not be too far above the saturation intensity, as that broadens the effective linewidth of the atom, which would smear the resulting spectrum a little. Now, for any atom in the sample where the probe beams are not on resonance, the pump beams do not scatter light, and so experience no attenuation. Therefore, the subtraction will show no signal. The same happens for any atoms for which the probe beam is resonant, but not the pump, as the pump is effectively not there. However, atoms that see *both* the pump and probe beams as resonant are pumped to saturation by the pump beam, and so scatter the probe beam *less* than they would in the absence of the pump. In these cases, one probe experiences less scattering than the other, and the subtraction of the two probe beams yields a signal. This happens when all beams are resonant with a transition of the atoms. In that case, the atoms that happen to have no motions along the beam axes will see all beams as resonant, and so the subtraction will yield a non-zero signal. Atoms that are moving, though, will only see one beam on resonance, and so will not contribute to the effect. Having an effect that only selects the atoms that are not moving cancels out the Doppler shifts, and the result is a Doppler-free spectrum.

There are a few caveats that should be mentioned. Rubidium atoms have multiple transitions available. While the different ground states have transition frequencies further apart than room-temperature Doppler shifts, the transitions from a single ground state to the various excited states are within that range. Therefore, if the pump and probe beams are tuned midway between two of these transitions, there is a set of atoms that are moving with just the right velocity along the beam axes such that the probe is resonant with one transition, and the pump is resonant with the other. The pump saturates its resonant transition, reducing the number of ground-state atoms that can absorb probe light on the probe's resonant transition. Thus, even if the pump and probe beams are *not* resonant with a transition, but are midway between two transitions (that are within the Doppler-broadened absorption spectrum), the subtraction of the two probe beams will show a difference. These extra peaks are called *crossover peaks*, and tend to be larger than the actual on-resonance peaks in the difference spectrum [66].

A further note on this setup is that the pump and probe beams do not need to be on the same frequency. If they are not, then we can transfer to a moving reference frame where they are of the same frequency. As long as the difference between the two frequencies is small compared to the room-temperature Doppler shifts, there will be atoms at rest in that frame which will see both beams on resonance at the same time. As a result of this, if the pump and probe beams are not at the same frequency, the difference spectrum will look as though the pump and probe were both at their average frequency. We typically lock to a peak in the spectrum so we know what frequency we are at; we usually use a crossover peak, because the signal is stronger. Elsewhere in the optical setup, we will use acousto-optic modulators to shift the frequency to the frequency we desire. In order to detect a peak, though, we need to dither the frequency by a small amount. If the difference spectrum changes as we shift the frequency, then we must not be at the peak, and we use that difference to feed back to the laser and shift it back towards resonance. For some of our beams, we simply dither the laser frequency. For the further-detuned barrier beam we used in our experiment, the dithering is small compared to the detuning and so has little effect. For the pumping barrier beam and the repumping MOT beam, which are supposed to be resonant, these small changes do not shift far enough from resonance to significantly decrease the pumping rate, which is all that matters. However, the MOT trapping beam, which is also used for absorption imaging, needs to be more stable. For absorption imaging, the beam needs to be right on resonance without dithering. If we dithered the laser, then because the absorption-imaging pulse is short compared to the our dither period, the images we take would each be at a random frequency within the dither range, which is undesireable. Furthermore, during MOT loading, the optimum cooling frequency is half a linewidth away from resonance. Shifting the frequency enough to detect the slope of the SAS spectrum requires a dither that is small but not insignificantly small compared to the linewidth. This dither would therefore be enough to reduce the trapping effectiveness of the MOT. For these reasons, instead of dithering the frequency of the whole MOT master laser, we use an acousto-optic modulator (discussed in Section III.9) to shift just the pump beam. To avoid shifting the location of the beam passing through the modulator, we use a double-pass configuration.

The actual setup we use is shown in Figure 3.2. A single beam is picked off from the main laser beam and attenuated. The two probe beams are picked off by a slide of uncoated glass (weak reflections off of the front and back surfaces), and pass through a glass cell with a dilute vapor of rubidium. The rest of the beam is brought around to the other side of the cell. If we are not dithering the entire laser, we double-pass the beam through an acousto-optic modulator, to avoid positional shifts of the beam as we change the frequency. That beam is then sent through rubidium vapor cell in the opposite direction of the probe beams. We can use the mirrors to align it such that it is quite close to one of the probe beams. By missing the probe beam by opposite (but small) amounts on either side of the cell, we can be reasonably certain that the two beams overlap for most of the cell. The two probe beams are then focused onto two photodetectors, and subtracted electronically to create the difference signal. The electronics allow for different gains for each beam, to account for slight intensity differences of the probe beams and slightly different detector sensitivities.

Slave Lasers

Each MOT we run requires only a few milliwatts of repumping beam power, at a maximum, but they require on the order of tens of milliwatts of trapping beam power. After accounting for various losses in the beam paths (and coupling into single-mode fibers), the master lasers do not produce enough power. Rather than have multiple master lasers operating at the same frequency, we chose to amplify them with slave lasers instead.

The slave lasers are simplified versions of the master lasers. They contain the same type of diode, but without a cavity. The diode is enclosed in a similar fashion

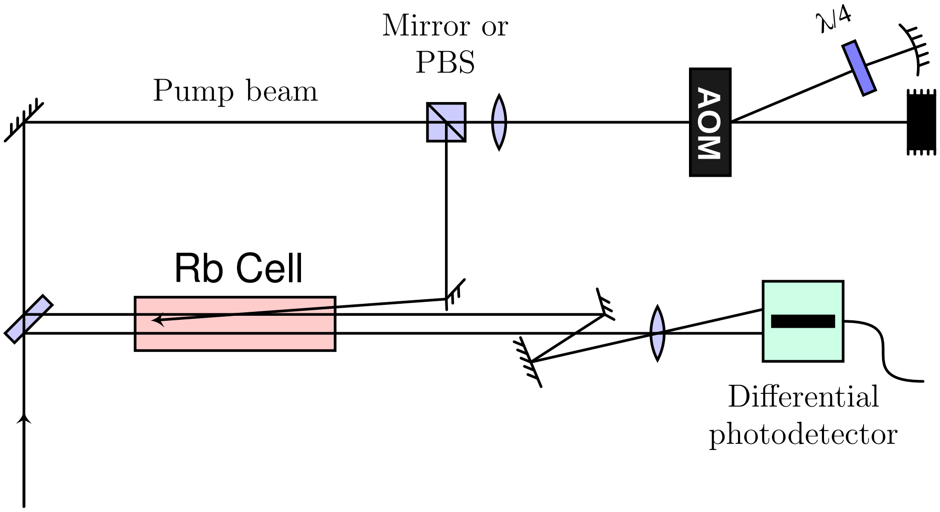


Figure 3.2. The basic schematic for saturation absorption spectroscopy. Two probe beams are picked off and pass through a cell of room-temperature rubidium vapor. The majority of the beam is used as the pump, and is passed through the cell in the reverse direction, crossing one of the probe beams. The two probes are focused onto a differential photodetector. Also shown is the optional double-passed acousto-optic modulator that we use to dither a beam if we do not want to dither the entire laser beam. In the setups without the double-passed modulator, the polarizing beam splitter (PBS) is replaced with a mirror, and the double-pass setup to the right is left out.

(but smaller), with temperature stabilization, some vibration isolation, and static protection. They also have the same series of optics in front of them, shown in Figure 3.3. A free-running laser would drift too much, but if seeded with a beam of a sufficient power, they will turn into optical amplifiers. This process is called *injection locking*, and results in an amplified beam with the same polarization and frequency as the seed, and a slightly larger linewidth, but that is not a limiting factor for us [76–79].

The slave lasers are set up to emit vertically polarized light, with the same anamorphic prism pairs to make the beam roughly circular. To ensure the best mode-matching possible, we try our best to shine the seed beam back along the exact same optical path, with the same polarization. To insert it into the path, we use the optical isolator. A beam shining backwards through the optical isolator will be rejected, unless we send it in through a rejection port.[4] Such a beam will then pass through the same optics as the slave beam, in reverse, and hit the slave with the same polarization that it would produce on its own. We can align the two beams on top of each other by eye, with the slave turned on. By marking where the slave beam exits the isolator, we can also try turning the slave off. Then, we can try to get the master seed beam to reflect off of the back of the slave diode and try to get the spot after the isolator to match up with where the slave beam was. This is sufficient for a rough alignment. For fine-tuning, we turn the slave on and shine it into a Fabry-Pérot cavity. By sweeping the master laser frequency a small amount, we can watch if the slave frequency sweeps as well; if it does, then the slave has locked to the master. We can then attenuate the seed beam until the slave loses lock, and tweak the seed beam alignment to try and recover the lock. To aid in this, the seed beam passes through an acousto-optic modulator to allow for easy attenuation. Improving the alignment makes for a lower power threshold required to lock the slave to the master; and lower that power threshold is, the stronger the lock will be once we turn the power up. We can also adjust the temperature and current of the slave lasers, although in practice we did not need to do as much of this as with the masters. We simply try to get the slave temperatures close to what the master laser temperatures are, and then run the slave lasers near the maximum output power.

The locked slave lasers can output over 100 mW, which is more than sufficient to run our trapping beams. The same master seeds two such lasers, one for each MOT. These beams are passed through an acousto-optic modulator that acts as both an optical switch and attenuator for the beam, and then coupled into a single mode fiber for transport to the MOTs. For the pyramid MOT, we couple the repumping and trapping beams into the same fiber, using a polarizing beam splitter to get the two beams on the same path (with orthogonal polarizations). For the six-beam MOT, the two beams are coupled into two separate fibers, which enter a two-to-six beam fiber splitter, which takes two input fibers, and splits their light roughly

[4] This process is described in slightly more detail in Section III.9.

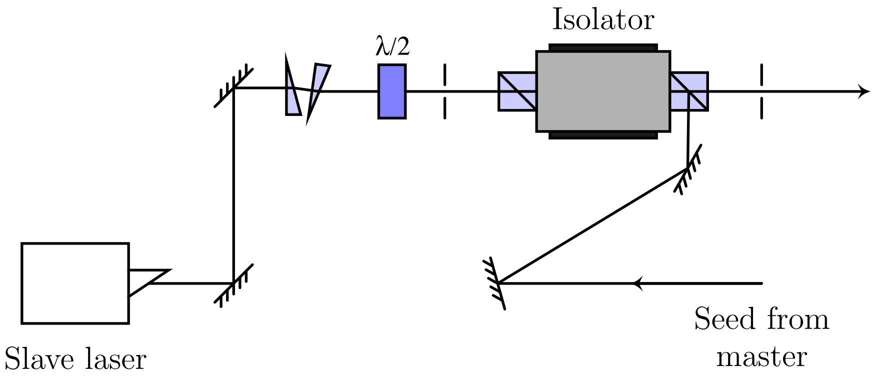


Figure 3.3. A schematic of the slave laser optics. The slaves have the same initial optics as the master laser shown in Figure 3.1. The only difference is we use a rejection port on the isolator to insert light from a master laser into the beam path, which seeds the slave.

equally across six output fibers. That gives us the six beams for the main MOT, each with a small amount of repumping beam power.

Magnetic Coils

Each of the MOTs has five separate magnetic coils associated with it. Aside from the anti-Helmholtz coils required to get the trap to work, we also have three pairs of Helmholtz coils.[5] The Helmholtz pairs are wired in series with each other, and so act as a single coil with a large gap between turns.

The pyramid MOT, in order to quickly collect atoms, is larger than the six-beam MOT. In order to operate, it also needs a large beam, which, in turn, requires larger coils to allow the beam to pass through. In a pyramid MOT, the initial incoming beam and the outgoing beam (after two reflections) have the opposite handedness compared to the perpendicular beams; this requires the anti-Helmholtz coils to have their axis aligned with the incoming beam, which itself is aligned with the atomic beam. This means that one of the coils has to encircle the narrow tube through

[5] They are not in the perfect Helmholtz configuration, as space constraints around the vacuum chamber prohibited this, but they are reasonably close.

which the atomic beam passes from the higher-pressure side of the chamber to the lower-pressure side of the chamber. As described in Section III.1, each side of the chamber has very large vacuum pumps attached to it which cannot be removed once we bake the chamber. Making the coils large enough to reach around the vacuum pumps is infeasible, and so we made the coils bake-able. We then assembled the chamber with one of the two anti-Helmholtz coils in place, and baked it.

To make a bake-able coil, we started with an aluminum coil form, which also aids in heat dissipation. We then wrapped the coil with 23 AWG copper magnet wire from MWS Wire Industries. We chose an insulation (Polyimide-ML) capable of being heated to 240°C (our final bake was to about 200°C). By mounting the coils an a rotating mount, we were able to wind the wire directly from the spool onto the coils, counting the turns as we went. To hold the wire in place, we added a high-temperature epoxy to the coils as we wound them. The epoxy (part number 353ND from EpoTek) is rated for temperatures from -55°C to 250°C [80]. Finally, to prevent the magnet wire from rubbing against the aluminum (which would eventually rub through the wire insulation), we made inserts out of Teflon where the wire leaves the coil form.

The pyramid-MOT anti-Helmholtz coils were designed to be as small as possible and still fit over the flanges to the vacuum chamber, so as to be able to get as close to the center of the chamber as possible. This necessitated an inner diameter of ~ 5.5 in, with a distance of 3.2 in between the coils. When the coils are further from the center, more current (or turns of wire) is required to get the same magnetic-field gradient at the center (the gradient determines the trap strength). We wound 400 turns around each coil. Given the quoted wire resistance of $20.4\,(3)\ \Omega$ per thousand feet, we expect longitudinal field gradient of about $8.7\,(8)$ G/cm with a power dissipation of $12\,(1)$ W at a current of 1 A. We got the first MOT operating before assembling the Helmholtz coils, and so used different currents in the anti-Helmholtz coils to shift the MOT near the center of the pyramid (compensating partly for the Earth's magnetic field, and any fringe fields from the chamber itself). Once we installed the Helmholtz coils, we continued to use the anti-Helmholtz coils for the axial shift. One coil continuously operates at -1.21 A, while the other operates at

0.98 A. In case the power dissipation would make the coils too hot, we machined a water channel into them by cutting a groove parallel to the channel holding the wires, and then welding a sleeve over it. The aluminum weld had a few leaks which we plugged by attaching an aspirator pump to the water inlets, and using that to suck the high-temperature epoxy into the few leaks. When the epoxy cured, the water channels appeared to be leak-free.

As it turns out, we have not needed to run water through the pyramid anti-Helmholtz coils. At their normal running currents of about 1 A, they run almost hot to the touch (near 40°C), but not hot enough to be a problem near the chamber. We installed a temperature sensor in them to monitor their temperatures.

The coils for the six-beam MOT had much fewer requirements. They sit outside of the Hellma cell, which has an outer dimension of 30 mm (as described in Section III.1). The Hellma cell is at the end of the chamber, and easily accessible for simple optical access. As such, the coils can easily be placed after the chamber is fully assembled and baked, and can be placed much closer to the MOT than in the primary MOT. This means they do not need to have nearly as many turns and current, and so need to dissipate much less power. Our intention was to not have eddy currents when we rapidly changed the current in the coils (mostly turning them on or off). We chose to make the coils out of black Delrin, which is non-conductive to reduce eddy currents, easily accessible, and easy to machine. As these coils do not need to be baked and do not dissipate enough power to become hot, we did not need to pick a high-temperature material. We used a similar mount for winding as the pyramid coils, winding 216 turns on each coil, using the same magnet wire and epoxy as the pyramid coils. At their typical operating current of 0.69 A, we expect a longitudinal field gradient of 17 G/cm and a power dissipation of 0.86 W. The coils do not get warm to the touch when operating.

The pyramid MOT coils are mounted to the table. We chose to mount the six-beam MOT coils to the chamber. This way, the arms holding the coils run along the edges of the Hellma cell. By not blocking part of a face of the Hellma cell, we maintain as much optical access to the MOT as possible.

Anti-helmholtz coils are often connected in series, so that the current through them is always the same. That way, inevitable fluctuations in the current cancel themselves at the trap center, and result in small fluctuations in the trapping force. If the MOT is not centered, or the fluctuations in the two coils are different, then the fluctuations cause small shifts in the zero-magnetic-field point. This shaking can heat atoms, and is a problem for groups try to form Bose-Einstein condensates (BEC). We do not intend to form a BEC in the near future, and do not plan to use a magnetic trap regardless.[6]Fluctuations are less important in a MOT which is highly dissipative anyways, and so we elected to have each coil in the anti-Helmholtz pairs be driven separately.

We designed and assembled a modified version of the current supply circuits that we use to power the diode lasers that could drive all of our coils. The circuits can be operated in a slave mode where they follow the current setting of a master circuit (with an adjustable scale factor); we use this to keep the two anti-Helmholtz coils running at similar currents. Most of the circuits in our lab work off of a ±15 V internal power supply, which we wanted to keep for this circuit. However, the resistance of the pyramid anti-Helmholtz coils requires they be driven with about 15 V at their operating currents, without accounting for a larger resistance as the coils heat up. We also wanted the circuit capable of using larger voltages to overcome the coil inductance and hence turn the coils on (or off) faster. To accomplish this, the output stage of the circuits has two op-amps in a push-pull setup that run off a separate 40 V power supply.[7] This allows them to drive current to the full power supply voltage in either direction. However, it also means we cannot directly sense the current, since the voltage of any sense element might be higher than the power supply of the rest of the circuit. The sensing is handled by measuring the voltage across a 1 Ω sense resistor in series with the current output. To prevent a voltage overload, the voltage at either side is divided three with a

[6] Completely by accident, we found that our setup *could* magnetically trap rubidium 87.

[7] The chips get quite hot, even with the large heat sink and fans that we use with them, and so we usually run the power supplies at 30 V. As an additional precaution, the output op-amps will shut down automatically if they overheat. The main circuit detects this and shuts down as well, either when the op-amps overheat, or if the coils overheat, provided we hook up a temperature sensor.

paired resistor bridge. The sense resistor is supposedly quite stable to changes in temperature, but the voltage dividers are not. Since the signal is small (due to a small sense resistor value), small changes in the values of these resistors change the calibration of the circuit; that is probably the biggest problem with the circuit.

We use our coil-driver circuits to drive all the Helmholtz and anti-Helmholtz coils. The resistor and capacitor values of the PID feedback loop in each circuit were individually tweaked to be able to shut the coils on and off rapidly with minimal ringing.

In addition to getting the MOTs operational, we also need to be able to shift the MOT locations around. An easy way to do this is to use additional coils in a (approximately) Helmholtz configuration. Applying a small constant magnetic field to the anti-Helmholtz coil fields shifts the zero-point around, which allows us to shift the MOT position. The same coils can also be used to cancel the Earth's magnetic field. Because we do not need really sensitive corrections over a wide area, and are mostly interested in shifting the MOT, we took a simple approach to the coils. The coils are made of a few turns of 8-conductor ribbon cable, connected together so that all the conductors are in series. This effectively multiplies the number of turns by 8. The coils are wrapped around a polycarbonate rectangular parallelepiped frame mounted on the table. This gives three pairs of coils. Each pair is hooked together in series and driven by a single coil circuit, for effectively three coils that shift the magnetic field at the origin in three orthogonal directions. In a true Helmholtz configuration, the first and second derivatives of the magnetic field at the center are zero; we only approximated that configuration, picking dimensions that could easily fit around the chamber and the existing optics.

Because the magnetic field gradients from the anti-Helmholtz coils are on the order of 10 G/cm, multi-millimeter shifts of the MOT require magnetic fields on the order of a Gauss. If the shift we want happens to be axial, imbalancing the anti-Helmholtz currents provides a much larger shift than we can easily achieve with the Helmholtz coils, which is one advantge to having the anti-Helmholtz coils independently driven, as opposed to connected in series. For this reason, and because we already knew the right settings before installing the Helmholtz coils, the pyramid MOT only uses the radial Helmholtz coils, and we use imbalanced currents in the anti-Helmholtz coils for the axial shift. For the main MOT, we keep the anti-Helmholtz coils balanced when the MOT is in the center of the cell, and use all three Helmholtz coils to keep it centered, and provide small shifts. For large vertical shifts, along the axis of the coils, we imbalance the anti-Helmholtz coils. We use such shifts when we load the MOT, as described in Section III.6. We used such shifts when we moved the MOT towards the bottom of the cell for our cooling experiments, which are discussed in the thesis of Elizabeth Schoene [37].

The MOT

Here we quickly overview our typical loading procedure for the MOT. For many of our experiments, we require the MOT to be about the same size for each repetition; therefore, it is important that the MOT loads for the same amount of time with each repetition.[8] To ensure this, we ensure all the atoms are dumped from the MOT by starting with all the beams off, and the six-beam MOT magnetic field coils off. We typically leave the pyramid MOT coils on. We then turn the six-beam MOT coils on and both the repump and trapping beams for both MOTs, with high intensities for better trapping. Normally, the six-beam MOT is below the atomic beam from the pyramid MOT. We found we could load the six-beam MOT much faster by imbalancing the anti-Helmholtz coils to shift it up more towards the atomic beam. We typically load for a few seconds; the actual loading time depends on the experiment and desired signal quality.

Once the loading is done, we move the MOT back to its normal location, and shut the beams to the pyramid MOT off. This turns off the atomic beam, which could introduce more background atoms to disrupt the six-beam MOT. After allowing 10 ms for the atoms to settle into their new location, we start the cooling phase. As described in Section II.2, polarization-gradient cooling works best at lower in-

[8] Alternatively, we could let it load long enough to reach steady-state. We elect not to do this because then individual repetitions would take too long to amass a data set with many repetitions. If repetitions and number of atoms are not important, we might not dump the MOT at all, and simply move it around.

tensities and larger detunings [56]. Therefore, we detune the MOT trapping beams by about 60 MHz, and cut the intensity of the beams by about a factor of four. We typically allow 20 ms for cooling, which cools the MOT from about 100 μK to about 30 μK.

Dipole Trap

A MOT is a very useful trap in itself, but the constant scattering of light required for its operation would destroy the types of experiments we wish to pursue, both ones that require very little interaction, and ones trying to demonstrate a non-Doppler method of cooling. We wanted to avoid using magnetic traps because they typically require large power supplies, the coils are usually hard to move, and we would need to be very careful of stray magnetic fields and nearby magnetic materials. Instead, we elected to use an optical dipole trap [6, 42, 57, 61].

The setup of our optical dipole trap is simple. We have a beam that exits a fiber, which is collimated, and then passed through a plano-convex lens with a 200 mm focal length. This focuses the beam down to a point near the MOT, which is all that is needed for the trap.

Initially, we attempted to use one of our master lasers tuned to about 784 nm for a dipole trap. This was passed through a tapered amplifier to achieve several hundred milliwatts of power, which we focused onto the MOT. This dipole trap performed miserably. We eventually guessed that the problem was the spontaneous emission background of the tapered amplifier. The spectrum of the background spans several nanometers, and so would include light resonant with rubidium 87. Having even a small amount of resonant and near-resonant light, focused tightly on confined atoms, would rapidly heat the atoms out of the trap. With that in mind, we triple-passed the beam through a cell of rubidium vapor heated to 80°C before coupling the light into the fiber. The rubidium in the cell would absorb (or, more accurately, scatter) resonant and near-resonant light. Heating the cell causes a larger fraction of the atoms in the cell to exist as a vapor, increasing the amount of absorption. This change took us from being unable to trap any atoms at all to being able to hold atoms for 100 ms, after which we could no longer detect any trapped atoms.

We then switched to a Ytterbium-doped fiber laser. The output is already fiber coupled, and terminates with a collimated beam. This outputs light at 1090 nm in a collimated beam with a 5 mm $1/e^2$ intensity radius. While the power output can range up to 20 W, for most of our trapping experiments we used 10 W. Coupled with the focusing lens for a dipole trap, this produces a trapping beam with a beam waist of about $30.9\,(5)\ \mu$m, and a Rayleigh length of $2.7\,(5)$ mm. A detailed calculation using far-off-resonant approximations (including counter-rotating terms) and multiple transitions suggest that this beam should present a trap for rubidium 87 atoms with a depth of about $k_{\mathrm{B}} \times 0.93$ mK, with a scattering rate on the order of 3/s.

The fiber output is mounted on a pair of translation stage with micrometers that let us shift the beam transverse to the beam direction. This entire setup is mounted on an Aerotech ABL10100–LT precision air-bearing translation stage. This stage is capable of positioning itself with sub-micron accuracy over a range of 100 mm. The micrometer translation stages are used only to align the beam onto the MOT. Having the trap on a computer-controlled translation stage allows us to move the atoms back and forth in the Hellma cell. The intent is once we are doing our quantum measurement experiments, we may want to trap the atoms, and then pull them back away from the MOT coil and optics for less encumbered optical access. During the cooling experiment in which we cooled atoms with out one-way barrier (described in Elizabeth Schoene's thesis [37]), we used this cart to translate the atoms. We selected this stage because another gro up found that a larger stage in the same product line could be used to successfully translate a Bose-Einstein condensate in a dipole trap [81]. While we do not need quite such precision, we figured it would not hurt.

Aligning the beam on the MOT is not too difficult, with the right tricks. By paying careful attention to where the beam entered and exited the cell, we could linearly interpolate to get it close to where the MOT was. We then backed the translation stage up so that we knew the focus was several Rayleigh lengths away from the MOT. This gave us a target that was much larger than the focus of the beams. By turning the beam on and off with a period of one to two seconds, we could watch the MOT on a real-time video camera. If the beam was hitting the MOT, it would shift the position of the trapped atoms. Once we had the beam overlapping the MOT, we started moving the focus closer to the MOT, using the micrometer translation stages to keep the beam on the MOT. Eventually, we got close enough that the beam waist was smaller than the trapped atom cloud. Since the security camera was aligned so that it viewed the atom cloud more aligned with the dipole trap beam than not, it could then see a hole in the atomic cloud where the ac-Stark shift from the beam shifted atoms out of resonance. We could then align this hole in the center of the MOT, and try to find the spot with the smallest hole, or at least the center of the region where the hole was small.

Once aligned, the exact same setup that just barely worked with the tapered amplifier at 784 nm worked great on the first try. The loading parameters took some optimization, and we eventually settled on the following. Once the MOT was loaded and cooled, we would turn the dipole trap beam on, overlapped with the MOT, and load for 110 ms. This loading time was partially based on a paper studying dipole-trap loading from Wieman's group [61]. They found that dipole traps load well for about 100 ms, but after that, interactions with the MOT beams caused excessive losses that started depleting the number of trapped atoms faster than they could load. We did a rough search, and found a similar optimum. During this loading, we detuned the MOT trapping beam further, and decreased the intensity of both the MOT trapping beams and the MOT repump beams, which we found loaded more atoms. Our main imaging camera is located to the side of the dipole trap, allowing it to image the full length of the dipole trap. We could then use that to align the focus of the trap (the center of the trapped atom cloud) with where we knew the MOT was from images of the MOT alone to further optimize alignment. We measured the lifetime of atoms in this trap to exceed 20 seconds.

Every so often, we would load the MOT and strobe the dipole trap beam slowly, to realign the beam with the MOT. Rather than shift the beam itself, we would make small adjustments to the magnetic field coils to locate the MOT, which could be done via a simple script on the controlling computer much easier than shifting the micrometer. As an additional benefit, not physically touching the setup made it much harder to bump something, and much easier to fix if we accidentally moved something too far.

The dipole trap beam can operate at up to 20 W, but even our normal operating power of 10 W can heat a surface up. If we let that power dissipate in our vacuum chamber, that could quite easily cause out-gassing where it heats the sides of the chamber. To avoid that, we built the chamber with a pair of mirrors in it. These mirrors deflect the dipole beam once it passes through the MOT, and redirect it through a window to the outside of the chamber. There, we have placed a large beam dump, which is essentially a heat sink with fins made of black anodized aluminum. The beam enters a long hole with a cone inside to aid in the beam absorption. If the dipole trap beam is on for long periods of time, this beam dump can get quite warm.

Imaging the Atoms

We use both absorption imaging and fluorescence imaging, described in Section II.4. Because it gets a stronger signal, we used fluorescence imaging for measurements that do not require much spatial resolution, such as determining atom counts and temperatures. For cases that do require spatial resolution, such as trying to determine what fraction of atoms are on each side of the one-way barrier, or finding the center of the atoms cloud in the dipole trap, we used absorption imaging.

In both cases, we use the same data camera, a MicroLine from Finger Lakes Instrumentation (FLI).[9] The cameras produced by FLI are intended for amateur astronomers, but we judged they would be a good starter camera for a cold atoms laboratory, since they are sensitive, easy to use, and cheap when compared to the data cameras often in use. We are quite happy with the MicroLine, although we

[9] This camera worked well for our barrier experiments, but we need a more sensitive camera for our next experiments. We are switching to a superior (and much more expensive) camera, as described in Chapter V.

wore out several of the shutter components years ago. The shutter inside the camera has four thin steel blades, each with two small ball-bearings spot-welded to it. One ball-bearing sits in a hole and acts as a pivot. The other, on the opposite side of the blade, fits in a groove on the shutter drive. The shutter drive is made out of a rather soft non-metallic compound, and was wearing down over time. For a period, we often needed to open the camera up and clean residue from the drive off of the shutter blades, which caused them to stick. Eventually, the ball-bearings started breaking off of the shutter blades. We welded them back in place, only to have them break off again. Eventually, we machined some some small rods of similar dimensions that fit in the holes where the ball-bearings had been, and welded those in place. They have not broken off again, and we have not needed to clean the shutter since.

Our MicroLine camera has a Kodak KAF-0402ME charge-coupled device (CCD) sensor, with a quantum efficiency of about 70% for 780 nm light. This is reduced slightly by a piece of Schott glass we installed in front of the sensor to reduce the signal from other light, such as the room lights. We previously had a laser-line filter in there as well, but took that out as it caused a strong etalon effect (interference fringes) with absorption imaging. With the lens we use for imaging (approximately two-to-one, with a one-to-one attachment that we do not use), a single pixel of the camera represents a square about 24.4 μm in the focal plane. The camera views the atoms in the dipole trap close to perpendicular to the dipole-trap axis.

For fluorescence imaging, we simply turn on the MOT beams (trapping and repumping beams) as if we were operating the MOT. The mechanical shutter on the camera is fairly repeatable, but only within a several milliseconds, and is highly dependent on how sticky the blades are. Since we often used a pulse of around 20 ms for fluorescence imaging, this could produce a huge fluctuation in the actual exposure time. Furthermore, the shutter took several milliseconds to open and close, so that different parts of the image had different exposure times. We solved this problem by leaving the beams off until we knew the shutter is fully open, and then strobed the MOT beams. While this a few milliseconds of dead time between the end of an experiment and the actual image, it gave us a well-defined imaging time that was consistent across the entire image.

The absorption imaging is a little more complicated. We use the same laser light used for the MOT trapping beam, with the frequency shifted to be on resonance. An acousto-optic modulator is used to shutter the beam, and to keep the intensity quite low. This beam is reflected off of the front surface of a 2° wedge of glass, through an anamorphic prism pair, and then through the Hellma cell directly into the camera. The wedge acts as a beam splitter and an attenuator; we chose a wedge because etalon effects caused large interference fringes in the images, and so we were trying to reduce interference everywhere we could. We also tried to make the absorption imaging beam slightly off-normal to the Hellma cell, and shifted it around so that the worst fringes, seen by the camera, did not overlap the area of interest in the image. The anamorphic prism pair spread the beam horizontally, so that it covered the entire extent of atoms trapped in the dipole trap (several millimeters) with enough intensity for a decent absorption image. The illumination was not very uniform (we show some sample images when we discuss our image processing in Figure 4.10), but as described in Section II.4, uniformity is not that important. We will cover the reason for a beam splitter shortly.

The imaging process was then to turn the repumping MOT beams on for about 0.1 ms, to ensure all the atoms were in the $F = 2$ ground state so that the absorption beam would be resonant. The time was judged short enough that atoms would not move much in that time. We would then flash the absorption beam for about 12 μs, while keeping the repumping MOT beams on. The camera would image this, with a faint shadow from the atoms. By repeating this without the atoms, we could subtract the two images and normalize to the image without atoms to see a snapshot of the atoms.

In order for this to work, it is imperative that the two images have the same intensity of illumination. One of our largest problems were the fringes in the image, which tended to change slowly over time. We were able to move most of the fringes from the Hellma cell away from our region of interest by making small changes to the angle of the camera and absorption imaging beam. There still remained a large circular fringe that would slowly change over time. After some investigation, we decided that fringe was most likely due to interference in the lens system, and shifted slightly as the temperature of the camera changed. Our best defense against this was to take take the repetitions fairly quickly, before the fringes could shift substantially. We also took many repetitions, which helped to average the effect out.

Our other main problem was that the intensity of the absorption beam varied from shot to shot. We eventually tracked this down to the AOM that shuttered the absorption beam. The AOM was driven with a low-intensity signal so as to produce a very weak first-order beam for imaging. At this low intensity, the AOM was quite sensitive to temperature. If simply turned on and left on, the first-order intensity would fluctuate, and eventually settle down. However, from shot to shot, the initial intensity varied too much for decent imaging. Our solution was that if we could not easily fix the intensity, we could fix the flux of the absorption beam, which is the important quantity. This why we have a beam splitter in the setup. The first reflection produces the imaging beam, while the transmission enters a photodiode. This photodiode is hooked up to a simple integrator circuit. We trigger the absorption beam AOM through this circuit, which integrates the intensity until it reaches a certain level, and then shuts the beam off. This resulted in very consistent images by varying the pulse length to counter the intensity changes. Over time, as the fiber coupling of the beam became worse, the pulse length would become much longer than the approximately 12 μs, which could be fixed by simply re-coupling the beam into the fiber again. Other than that, the setup does not require much effort.

Extra Hardware

In this section, we will briefly describe some of the non-standard hardware we use in our lab, as well as quickly cover the ways in which we use some standard hardware.

We start with acousto-optic modulators, a fairly standard piece of equipment in a physics laboratory. These are basically a crystal attached to a high-frequency piezo actuator. The piezo is driven with a strong signal at a high frequency, setting up a sound wave in the crystal. This sound wave manifests as a time-dependent density wave, which in seen by a laser beam passing through the crystal as a time-dependent index of refraction. The period of the wave is short enough to diffract the beam; the time-dependence modulates the beam as well, causing a shift in frequency. The result is that part of the beam passes through the crystal (the zero-order beam), while other parts get diffracted a small amount to the side (the ±1 first-order diffractions). The +1 diffraction frequency is up-shifted by the frequency of the piezo signal, while the −1 diffraction beam is down-shifted. There are higher-order diffractions as well, but we typically only use one of the first-order beams. By fine tuning the angle at which the beam hits the crystal small amounts away from normal, we can actually emphasize the +1 first-order diffraction peak over any other.

Because the piezo operates at such a high frequency, we can easily turn the signal to the piezo on and off with sub-microsecond timing. This, in turn, allows us to turn the first-order beam on and off with the same timing. Furthermore, by adjusting the amplitude and frequency of the piezo signal, we can adjust the amplitude and frequency of the first-order beam. This allows us to use acousto-optic modulators as very fast optical switches, attenuators, and frequency shifters. We use the first-order beam for the AOMs, which is typically shifted up by 80 MHz (the frequency we use for the opctical-switch AOMs), because the zero-order is never attenuated very well, while the first-order is completely extinguished when the AOM is turned off. The only catch is shifting the frequency of the piezo changes the period of the diffraction grating, which shifts the angle of the first-order beam. In cases where this shift is problematic, we double-pass the beam through the modulator. To accomplish this, the beam passes through a polarizing beam-splitter cube, and lens with a long focal length. The modulator is placed at the focal point of the lens. The first-order diffraction beam hits a curved mirror with a radius of curvature equal to the focal length of the lens, such that the modulator is at the center of that curvature. This way, the first-order beam, regardless of exit angle, is back-reflected through the AOM a second time. The lens and mirror combination cancel focusing/spreading effects of the other, and the result after passing the lens a second time is a collimated beam in the other direction. With the addition of a quarter-wave plate somewhere in the beam path, the two passes make it effectively a half-wave plate, arranged to rotate the polarization of the beam by 90°. When the beam hits the polarizing beam-splitter cube, having a different polarization means it takes a different path from the original beam, and we have a second beam with a different frequency that does not shift when we change the frequency.

Optical isolators are another standard piece of equipment. We briefly describe them here to explain how we seed our slave lasers, as described in Section III.4. Ordinarily, Maxwell's equations are invariant under time-reversal. The one exception is magnetic fields, where a reversal of time reverses the currents that generates the fields, and hence reverse the fields. By taking advantage of this asymmetry, isolators use a crystal under a strong magnetic field to break time-reversal symmetry in a beam. The result is a crystal that rotates a linear polarization in a right-handed manner for a beam traveling in one direction, but a left-handed manner for a beam traveling in the other. Put another way, for an observer looking down one axis, the polarizations always rotate clockwise (or counterclockwise, if the observer switches sides), regardless of beam direction. The isolators we use have a polarizing beam-splitter cube at either end, one set to transmit a 45° polarization, and the other set to transmit a vertical polarization. We send a beam in the 45° polarization side, and the crystal rotates if 45° so that the outgoing polarization is vertical, which transmits through the second beam-splitter. Any beam that gets back-reflected passes through the vertical polarizing beam-splitter, and the polarization gets rotated in the *same* direction, and the end result is a $-45°$ polarization which is *reflected* by the polarizing beam-splitter cube. This reflection is usually blocked.

To seed our slave lasers, the isolators are set up to output horizontally-polarized light instead of horizontally-polarized light. We send a vertically-polarized beam from a master laser into the side of the horizontally-passing beam-splitter cube. If aligned properly, this beam reflects into the main beam path through the isolator. As it heads backwards, the same rotation that would change a horizontal polarizeation into a $-45°$ polarization changes a vertical polarization into a $+45°$ polarization, which transmits through the beam-splitter. This beam now back-propagates along the original beam path, with the same polarization. This is how we seed our slave lasers.

We also developed a really cheap beam profiler to help us measure the beam waists and locations of the beams in our one-way barrier. Beam profilers are essentially just imaging detectors, often with specialized software. We basically took a cheap webcam intended to use in the home for watching pets, called the PetCam Network Camera, made by Panasonic. Cheap webcams such as this employ detectors that are fairly sensitive in the visible and near infrared, which means they work great for the wavelengths we use. Furthermore, the only software required to use one is a web browser, which we already had. More specifically, they only require software that can download an image via Hypertext Transport Protocol (HTTP); as there are many scriptable programs and libraries that can do this, it was easy for us to roll our own automated library to take images and curve-fit them. The pixels of the sensor are 5.6 μm on a side, which allows for decent resolution. Naturally, the webcam is intended to take images, and so has a lens. Removing the cover revealed that the lens was mounted in a plastic holder directly over the sensor. We were able to remove the lens with a razor, and then cut a hole in the front of the case to glue a neutral density filter holder in place. This let us attenuate the laser beams we were profiling, as even weak laser beams can quickly saturate these detectors.

With a browser, we can easily watch the beams in near real time while we adjust beam intensities or location in search of a particular focus, and then run an automated script to take data. The connection is over a relatively cheap Ethernet cable, which can be purchased or assembled with little difficulty. As most of our computer-controlled equipment is controlled via an Ethernet network, we already had the materials and the network to operate the camera. The only downside is the image is compressed in a fairly lossy fashion. We could mitigate this by averaging over many images to get a decent image. Some sample images are shown in Section IV.3, Figure 4.6.

We conclude with a quick discussion of some of the various mechanical shutters we have used in our lab. We use acousto-optic modulators for the majority our our shutters, but they do have some leakage when they are off, so it is handy to back them up with a mechanical shutter. Our first shutters were some electronic relays, to which we glued a thin arm with part of a razor attached to the end [82]. By triggering the relay, the arm would move, and the razor could be used to shutter a beam. The downside to these relays is they are rather weak. Having such a long arm tends to cause the arm to bounce when it hits the stop. When we added a small amount of sorbothane to damp the bouncing, the arm tended to stick a little. This, combined with a small deformation of the sorbothane, made the shutters unreliable. If they were too weak, they might not shut at all; if they were too strong, they tended to bounce. Either way, the time between the signal to when the beam turned off had a little too much hysteresis.

The cheap shutter design that we do use on our table is a small speaker [83]. By cutting the paper cone off the speaker, we got direct access to the voice coil. We simply cut a simple form out of aluminum that holds the magnetic assembly in place, and keeps the voice coil from jumping out of the groove in which is slides. This form also allows us to mount the speaker on the optical table. By gluing a piece of black plastic or metal to the shutter, and placing it at the focal point of a one-to-one telescope, we found we could completely block (or unblock) the beam with sub-millisecond precision, which was quite adequate for our uses. The typical delay between signal and actual blockage of the beam was a few milliseconds (with a standard deviation of a few tenths of a millisecond), and the actual blocking time was smaller still than that. Since these are speakers, the shutters can be cycled rapidly. The only downside for these speakers is the copper braids attached to the voice coil can wear out and break. However, after soldering them a few times, we eventually got them to be stiff enough that they stopped wearing.

Our final attempts at a rapid shutter was to disassemble an old hard drive [84]. By replacing the read-write head, which is on a movable arm, with a piece of a razor blade, we could make a shutter. The shutter can be controlled by wiring directly into the actuator that moves the arm. Others have used this design, but by the time we tried this, we already had our speaker shutters working, and so did not work a great deal with these shutters. The large arm tended to make the shutters much noisier than our speaker shutters.

The Overall Setup

In this section, we show how our setup is arranged. The vacuum chamber was described in Section III.1. Due to the need to bake the chamber, we tried to separate the optics and the chamber as much as possible. To this end, we have a somewhat minimal set of optics near the vacuum chamber, and all the light is brought to the chamber through single-mode optical fibers. The setup for the barrier beams, described in Section IV.3, was kept on a moveable breadboard. The MOT beams come from fiber couplers with mounted quarter-waveplates. We placed a pair of crossed rails on the table to help us align the beams for the MOT. By printing cards with marks at the right height, we could place the cards on the rails and shine the beams through the holes to line them up with the MOT. The holes in the cards are slightly shifted to account for the shift as the beams pass through the Hellma cell.

On the optics side of our table, we have four master lasers. One of the lasers drives a tapered amplifier. We had intended to use this laser as a dipole trap, but, as described in Section III.7, this failed. We may convert this setup into an additional trapping laser, but for now, it is unused. The remaining three master lasers provide the repumping light used in the MOTs and the one-way barrier (the pumping beam), the MOT trapping light and imaging light, and the barrier beam light.

The MOTs do not require a lot of repumping light, and we found a single master laser was sufficient to provide all the light we use. The schematic of the light is shown in Figure 3.4. Light from the master laser passes through the usual initial optics, described in Figure 3.1. We then pick off two beams. One of the beams goes to a saturated absorption spectroscopy setup, described in Section III.3 and detailed in Figure 3.2, that we use to lock the laser to resonance. As the repump beam may be shifted small amounts from resonance and still be able to pump atoms adequately fast, we chose to dither the frequency of the main laser by modulating the drive current as opposed to adding a double-passed AOM to the saturated absorption spectroscopy setup. The other beam is coupled into an optical fiber as a spare beam, which we use as a wavemeter reference. The main part of the beam passes through an acousto-optic modulator that we use to turn the repump beam to the MOTs on and off (the same AOM controls the light to both MOTs). The first-order diffraction is passed through a half-wave plate to rotate the polarization to 45° and a one-to-one telescope. The telescope helps ensure the beam has the right divergence to couple into the single-mode fibers. Having a 45° polarization allows us to split it into two beam with a polarizing beam-splitter cube; the reflected beam is coupled into a fiber for the pyramid MOT, and the transmitted beam is coupled into a fiber for the six-beam MOT. The fiber for the pyramid MOT has a second polarizing beam-splitter cube to allow us to couple the trapping light in as well; at the end of the fiber, the light exits without a coupler, so that it spreads into a large beam, passes through a quarter-waveplate, and then a large lens to collimate it and form the pyramid MOT beam.

The zero-order in the repumping setup is coupled into another fiber. That fiber exits onto another acousto-optic modulator; the zero-order there is dumped, and the first-order is coupled back into a fiber that provides the pumping beam of the one-way barrier. We never needed the pumping beam of the barrier and the repumping beam of the MOTs to be on at the same time, so we used the second AOM to enable the pumping beam (and control its intensity) when the repumping beams were off.

The MOT trapping is the most complicated of the laser setups because it needed to be amplified for each MOT. A schematic is shown in Figure 3.5. Initially, the beam passes through the usual set of optics, described in Figure 3.1. Part of the beam is picked off for the saturated absorption spectroscopy, detailed in Figure 3.2. In this case, we did not want to modulate the entire laser light, and so the pumping beam has an extra double-passed acousto-optic modulator. This modulates the pump beam only, which provides the dither that we use to lock the laser without modulating the entire laser beam.

The main part of the beam not used for locking the laser passes through another double-passed acousto-optic modulator, which we use to vary the frequency of the MOT beam. This is then passed through a half-wave plate that rotates the polarization to 45°, utilizing the same polarizing beam-splitter cube to do split the resulting

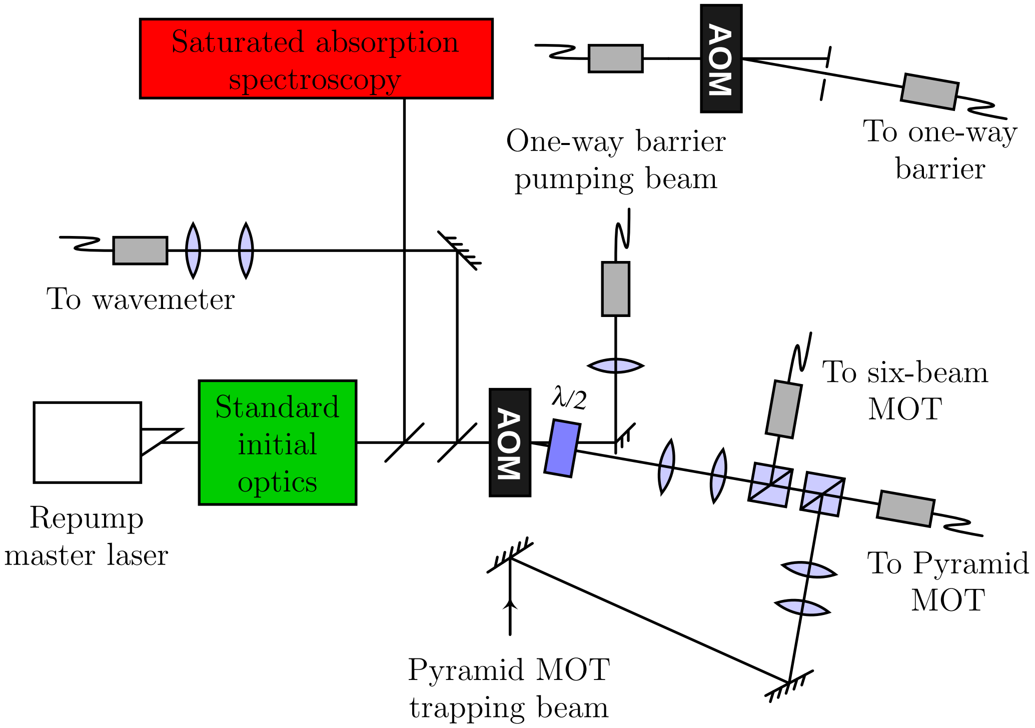


Figure 3.4. A schematic of our MOT repumping beam. The standard initial optics are shown in Figure 3.1, while the saturated absorption spectroscopy setup is shown in Figure 3.2. The repump beam dithers the entire laser, and so does not use the double-passed AOM in the saturated absorption spectroscopy setup. The one-way barrier pumping beam fiber connects to the setup in the upper-right. That beam passes through an AOM, and then couples into another fiber which connects to the barrier beam table, shown later in Figure 4.4.

beam. Both the beams are aimed into the rejection ports of the isolators for slave lasers; the reflected beam seeds the pyramid MOT slave, and the transmitted beam seeds the six-beam MOT slave.

The pyramid MOT slave light that leaves the isolator passes through an acousto-optic modulator. The zero-order beam is passed to a Fabry-Pérot cavity that allows us to check that the slave is locked to the master laser. The first-order beam is passed through a one-to-one mirror and coupled into the same pyramid MOT fiber used by the repumping part of the beam. The two beams are combined using a polarizing beam-splitter cube.

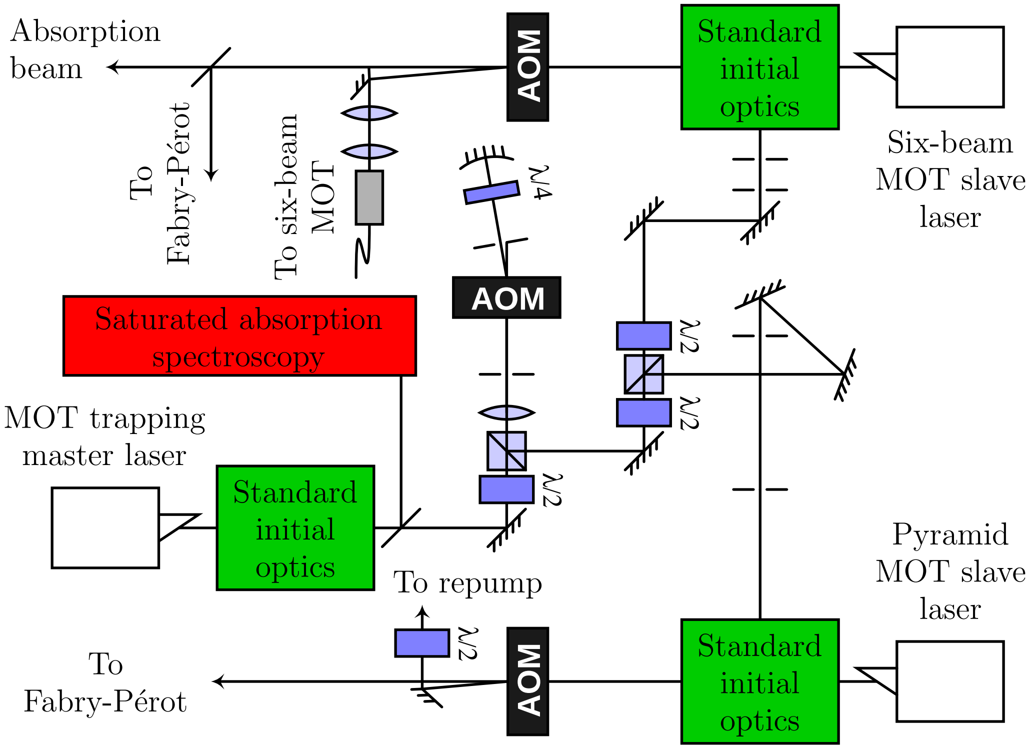


Figure 3.5. A schematic of our MOT trapping beam. The standard initial optics are shown in Figure 3.1, while the saturated absorption spectroscopy setup is shown in Figure 3.2. We use the double-passed AOM shown in the saturated spectroscopy setup for this laser. Both slave lasers are coupled into the same Fabry-Pérot cavity (not shown). The pyramid slave beam is combined with the repump beam as shown in Figure 3.4. The absorption beam is passed through an AOM in the same manner as the AOMs on the slaves shown here, with the first-order beam coupled into a fiber, which provides the beam we use for absorption imaging.

The six-beam MOT slave light leaves the isolator and also passes through an acousto-optic modulator. We let the beams travel a longer distance than usual to try and separate the two orders as much as possible. The zero-order is coupled into the same Fabry-Pérot cavity used for the pyramid MOT slave. A mirror on a flippable mount lets us select which beam we can see in the cavity. The first-order beam passes through the usual one-to-one telescope and is coupled into a fiber for the six-beam MOT.

Both of the MOT beams have a relay shutter (described in Section III.9) at the foci of their one-to-one telescopes. Initially we used those shutters to help block the beams when we had them off. Eventually we decided they were not worth the hassle, and those shutters are usually left open.

The two six-beam MOT fibers (repumping beam and trapping beam) enter a two-to-six fiber splitter that combines the light from the two fibers and splits it more-or-less evenly to six output fibers. The fiber outputs match to within about 10%. To prevent the variations from being a problem, we arranged the fibers so the pairs that were the closest matched in power opposed one another in the six-beam MOT. Each of the fiber outputs has a quarter-waveplate mounted on it to make the beam polarization the proper handedness for the MOT.

The final master laser powers the main barrier beam in our one-way barrier setup. This beam was the last one to be set up, and was already blocked in by the MOT trapping beam setup and the tapered amplifier setup, and so the beam path is relatively tightly wound, but the setup is the same. As with the other beams, the laser light passes through the usual set of optics shown in Figure 3.1, and part is picked off for the saturation absorption spectroscopy setup shown in Figure 3.2. As with the repump beam, we dither the laser current for locking, as opposed to having a double-passed AOM in the saturated absorption spectroscopy setup. The main part of the beam passes through a one-to-one telescope and is coupled into a fiber that takes the light to the barrier table shown in Figure 4.4. We installed a speaker shutter (described in Section III.9) in the focus of the telescope, which we used to turn the beam on and off.

We have described the basic beam paths used, and provided figures showing the layout of the table, but have not described the particular part numbers or relative powers at each part in the setup, nor certain alignment procedures. Those details have been covered in previous theses by Elizabeth Schoene and Tao Li, and we have omitted them here [36, 37].

Control Circuits

Our experiments require a large number of signals to be controlled with a high degree of precision. As a quick example, a typical experiment involving loading the MOT, possibly for 10 s, and ending with absorption imaging which consists of a 4 μs pulse to trigger the circuit that controls the absorption pulse. This pulse must also coincide with a pulse to trigger the repumping beam 0.1 ms before the pulse, to make sure the atoms are in the correct state.[10] That requires signals to be aligned with a precision better than 0.1 ms out of 10 s, or one part in 10^5. Furthermore, we need to be able to trigger the camera, two barrier beams, five coils, two sets of MOT beams, not to mention various intensities and frequencies.

With the aid of Peter Gaskell, who worked with our lab for a few years, our lab developed a precise but inexpensive laboratory control system [34, 35]. This system is based in part on circuits developed by Todd Meyrath and Florian Schreck when they were graduate students under Mark Raizen at the University of Texas, Austin [85]. These circuits come in several types; we use one that has sixteen digital (on/off TTL) outputs and another that has eight digital-to-analog converter channels, each capable of being independently set between ±10 V with sub-millivolt precision. The circuits can be daisy-chained together with a 50-pin ribbon cable, with several lines dedicated to addressing particular boards of the set.

Peter Gaskell designed an interface board that stores commands from an embedded Ethernet-enabled system called the Ethernut, and sends those commands over the 50-pin bus to the output circuits. The Ethernut takes a series of compressed commands over an Ethernet network, translates them to set commands for the output boards, and streams them to the interface board, attempting to keep the buffers on the interface board full. The interface board sends these commands to the output boards on a clock signal. For our output clock, we use an old rubidium atomic clock that we managed to acquire, making our clock perhaps more precise than our electronics. Each Ethernut/interface combination drives two separate output boards in our setup, although the hardware is flexible enough for other configurations. As long as the same clock signals are sent to every set of boards, and they all trigger off of the same quick pulse, the boards will all work in series to within the precision of the atomic clock or the limits of the electronic components. Since

[10] This procedure is described in Section III.8.

the system uses Ethernet cabling, with an Ethernet switch, we can easily connect a great number of these systems to a single computer without any extra hardware (assuming the computer has an Ethernet port). Ethernet networks are fast enough that we can send sequences to the boards rapidly. We therefore have an easily expandable number of analog and digital outputs that are easily synchronized, with every experimental procedure easily repeatable, to a high degree of accuracy.

To run the system from the computer side, we developed a library written in Perl that takes a series of events, converts them to programs to send to the Ethernut/interface boards, and programs the boards in parallel. Since Perl is a readily available cross-platform language, and the boards and equipment we use for the circuits are relatively inexpensive compared to various commercial alternatives, we believe that any laboratory could replicate our control system without much difficulty or expense. We also consider our setup to be easily modifiable, so that others may tailor the system for their own requirements.

We published a more extensive description of the system, along with some performance reports, in Review of Scientific Instruments [34]. We also keep an official web page where we store the schematics, drawings, and software [35]. As a final test of the system, most of the experiments reported in this thesis were accomplished using this system in some stage of its development. Not only that, but the system is easy enough to use and script that most simple tasks, such as enabling the MOT, or turning various beams on or off, are automated through this system rather than toggle various hardware manually.

CHAPTER IV

AN ALL-OPTICAL ONE-WAY BARRIER

Our first big experiment as a lab was the creation and testing of an all-optical one-way barrier for neutral rubidium 87 atoms. The basic theory for our particular experiment was worked out by Mark Raizen's group [16], although there was a similar proposal from Ruschhaupt and Muga [23]. We thought the one-way barrier would be a simple way to test the laboratory equipment and setup we had just recently finished developing and assembling. The technique turns out to have had more difficulties hidden in the details than we had anticipated, and what had been intended as a simple test turned into a rather involved, but successful, experiment [18, 20, 21].

This chapter presents an overview of how we set up and tested our one-way barrier, along with some computations and simulations concerning it. Two other theses from our lab provide some complementary explanations and more details on the robustness of the one-way barrier [36, 37].

Uses of a Barrier

An atomic-level one-way barrier can be thought of in terms of a more macroscopic one-way barrier. At the human scale, we can imagine several types of one-way barriers. Some, such as the retractable spikes in some parking spaces intended to keep cars from driving out the entrance, are true one-way barriers. Others, such as a parking gate operated by a parking attendant, work on more of a decision-based principle.

In the case of the retractable spikes, or any ratchet-based system, there really is a physical asymmetry that makes the system uni-directional, even for systems that work at the molecular level [86, 87]. Oddly enough, these sorts of systems require some sort of underlying complexity to work. The rules that govern single atoms dictate the same motions going backwards in time as forwards, making it difficult to create a system where an atom can pass a region going one direction, but not the other. Such systems require some sort of dissipation, where atoms collide with other objects, which in turn collide with yet others, until the effect is so spread out that it is unlikely to ever come back to the atom. The more decision-based one-way barriers, as we will discuss in Section IV.6, must reduce to this same sort of complex system at some level.

We could imagine a rather simple one-way barrier that relies heavily on such dissipation. Imagine containing atoms in a focused red-detuned laser beam, as described in Section III.7. We could then illuminate these atoms with some sort of near-resonant, red-detuned light, creating an optical molasses that slows the atoms down. If we crossed the focus of the dipole trap with a sheet of near-resonant light at an angle to the axis of the trap, that sheet would preferentially accelerate the atoms to one side of the trap. If we set up the experimental parameters correctly, the optical molasses would slow the atoms down enough that they could no longer cross that beam going against the flow, and we would effectively have a one-way barrier.

The intent of the barrier as we create it is to interact with the atoms as little as possible, using effectively conservative potentials. Dissipation is required for such a barrier to work, as described in Section IV.6, but dissipation requires some sort of interaction. Our goal is to have a one-way barrier where dissipation is only required when crossing the barrier, so that, once trapped by the barrier, essentially no further interaction is required. This way, the electronic state of the atom would not be constantly changing, and there could be some hope of having some sort of internal coherence between atomic states.

One advantage of such a scheme is in the context of cooling atoms [16, 17, 21, 22, 26]. Optical molasses, as easy-to-implement as it is, has some deficiencies. For instance, in most real-world atoms, there are states the atom could be pumped into that would not be cooled by molasses, requiring extra "repump" lasers. As an atom becomes more complicated, the number of repumping frequencies required increases, and the problem quickly becomes more difficult than useful. Also, the cooling limit for optical molasses, the Doppler limit, exists because the constant scattering of light off the atoms through which optical molasses works also heats the atoms. The Doppler limit is the point where the cooling effect balances the heating effect, optimized so that canceling occurs at the lowest possible temperature.

Any cooling scheme that involves less disruptive interactions with an atom or other object could overcome these limitations of optical molasses. The one-way barrier we demonstrated, along with similar realizations from the Raizen lab at The University of Texas, could, in theory, be used to cool atoms and/or molecules where the transitions are either too numerous or too inaccessible for optical molasses to be feasible [17, 19]. A similar scheme could also be used, theoretically, to cool below the Doppler limit. As our experimental setup was designed around our future experiments testing quantum measurement and quantum feedback, we demonstrated that the barrier could cool atoms, did not attempt to optimize our barrier for cooling. While the cooling project is not discussed in this thesis, the results are published elsewhere [20, 21, 37]. The publications include a discussion of how our method was limited by insufficient detunings and an apparent heating effect, and proposals on how to circumvent the limitations.

Optical One-Way Barrier Theory

To describe our one-way barrier with very little dissipation, we use the parking-gate attendant description. The gate is not inherently one-way; it is either open or closed. When a car pulls up, the attendant first checks which side of the gate the car is on. If the car is on one side, the attendant opens the gate to let the car through, and if the car is on the other, the attendant keeps the gate closed, and the car cannot pass. Even though the gate itself can let cars through in either direction, with proper control it acts as a one-way barrier.

This sort of gate could work at the single-atom level as well, but there is another problem. What if the cars (or atoms) were so dense, and the gate so big, that there are always cars on both side of the gate, trying to pass through? In that scenario, whenever the attendant opens the gate to let cars through, they will pass in both directions, and the gate will no longer be a one-way barrier. This is the case for most cold-atom traps, where the number of atoms is often in the thousands or millions. With such large numbers, one cannot wait for large imbalances of atoms trying to pass the barrier, because, while such conditions are bound to happen, the expected time required before such a condition would be likely becomes unfathomly large with a large number of atoms, especially once there are fewer on the side from which atoms are allowed to pass. This is further complicated by the fact that the atoms are often in a very small volume, where it is very hard to count how many atoms are on one side of the barrier versus the other.

The trick is to remove the gate attendant, and instead have each individual atom be its own attendant. We can effect the same one-way behavior by encoding *in each atom* whether they are allowed to pass through gate. If we then allow the atoms on one side only to pass, but not the atoms on the other side, we have our one-way barrier. We can avoid further interactions with the atoms by simply tagging the atoms that pass through one direction as no longer being able to pass through the gate.

We perform this encoding using the electronic state of the atoms. While a two-level atom with a sufficiently long lifetime in either state is sufficient, we use a three-level scheme [16]. Since the three-level atom theory is nearly identical to the two-level atom version, we only describe the three-level atom theory here.

In our three-level atom, we assume there are two long-lived "ground states"[1] coupled to a single excited state. In rubidium 87, we actually use manifolds (collections of atomic states) instead of actual single states, but the idea is the same. All that is needed is that the frequency difference between these two ground-to-excited transitions be large compared to the linewidth of the transitions, so that the frequency that excites one transition will not excite the other.

Figure 4.1 shows the energy levels and transitions we use for our barrier. Note

[1] Technically, only the lowest energy state should be called the ground state. However, in rubidium 87, these two "ground states" have lifetimes much longer than a year. Since an atom that decays into one of these two states is essentially done decaying, it is as though both are lowest-energy states, even though one is technically a little lower than the other, so we call them both ground states.

how an atom would be affected by a beam of light tuned between the two transition frequencies. If the atom is in lower of the two ground states, it requires a larger energy to reach the excited state, and so the corresponding transition frequency is higher, than if the atom is in the upper of the two ground states. Thus, an atom in the lower of the two ground states sees the intermediate frequency as below the transition frequency, or red-detuned, while an atom in the upper of the two ground states sees the intermediate frequency as above the transition frequency, or blue-detuned. If the detuning is large enough that it creates a quasi-conservative potential for the atom in the same manner as an optical dipole trap, then a sheet of light at this intermediate frequency will present an attractive well to atoms in the lower ground state (because they see the light as red-detuned), and the same sheet of light will present a repulsive barrier to atoms in the upper ground state (because they see the light as blue-detuned). Atoms in the lower ground state will be attracted into the well; however, in the absence of dissipation, they will gain enough kinetic energy entering the well to come out the other side of the well, which is essentially the definition of a conservative potential. Atoms in the upper ground state will be repelled from the barrier. As long as the beam is intense enough that the barrier height is greater than the kinetic energy of the atoms, this beam will reflect atoms in the upper ground state, while atoms in the lower ground state will pass through with their final kinetic energy unchanged. This intermediate frequency produces the selective potential that is only a barrier on atoms in a certain state, as shown in Figure 4.2.

We now have a way to encode in each individual atom whether the atom can pass through the barrier made by this intermediately tuned sheet of light, by selecting which ground state the atoms are in. We can now refer to the upper ground state as the *reflecting state*, and the lower ground state as the *transmitting state*. We also refer to this intermediately tuned sheet of light as the *barrier*, even though it is only selectively a barrier.

The final step to changing a selective barrier into a one-way barrier is to place the atoms in the reflecting state on only one side of the barrier, which we typically show as the right-hand side of the barrier. We do this with a second sheet of light

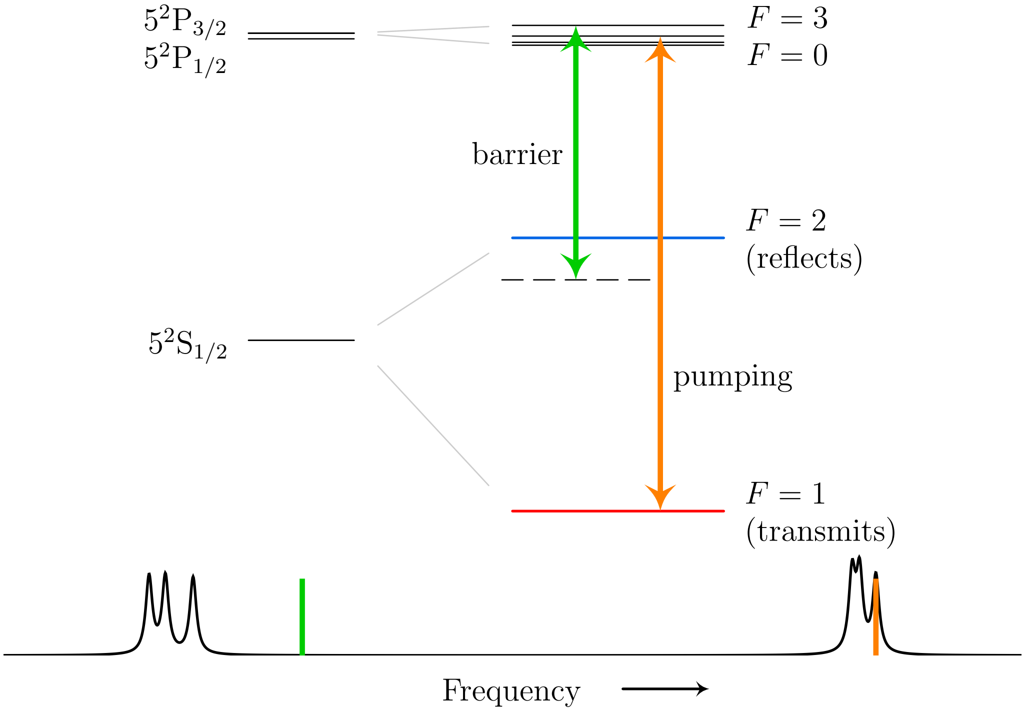


Figure 4.1. The frequencies we use for our one-way barrier. We show the ground-state manifold and the two excited state manifolds, with the upper one corresponding to the D_2 transition. On this scale, the hyperfine structure cannot be seen, so we magnify the upper excited-state manifold and the ground-state manifold by a factor of $50,000$. Below the levels, we show the spectrum corresponding to these levels. The set of peaks to the left correspond to the transitions from the reflecting state, while the set of peaks to the right correspond to the transitions from the transmitting state. As can be seen in this spectrum, the ground states are split by much more than the excited states and the linewidths, which is why we can approximate this as a three-level atom. We also show the frequencies for the barrier beam and the pumping beam, on both the level diagram and the spectrum. The pumping beam is resonant with a transition from the transmitting state, while the barrier beam is tuned between the two transitions, so that it appears blue-detuned to atoms in the reflecting state, and red-detuned to atoms in the transmitting state. The asymmetric detuning is intentional, for we describe later.

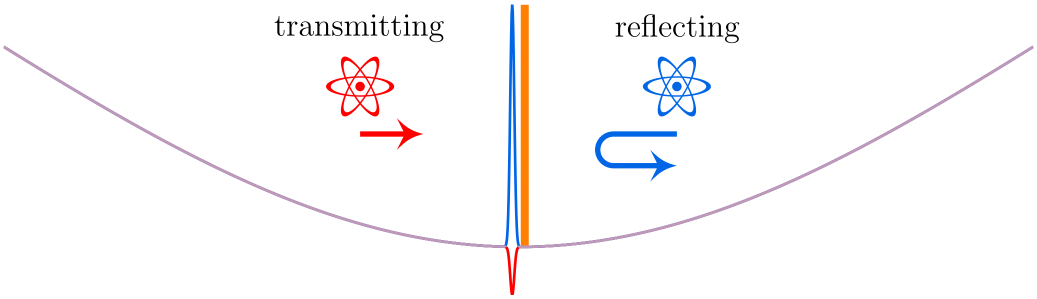


Figure 4.2. A simple schematic of our one-way barrier for three-level atoms. The atoms are all confined in a trapping potential, with the barrier beam in the center, and a pumping beam to one side. Initially, all the atoms are in the transmitting state, and so can pass through the barrier. Once they pass through the pumping beam, they are pumped into the reflective state, and become trapped on that side of the barrier. After that, they see the barrier beam as a strong repulsive potential which traps them on the right-hand side of the barrier.

that is resonant with the transition from the transmitting state to the excited state, as shown in Figure 4.1. This *pumping beam* will repeatedly excite atoms from the transmitting state to the excited state and back. While in the excited state, there is a chance the atom will decay into either of the two ground states. If it decays into the transmitting state, the pumping beam will continue pumping the atom. Eventually, it will decay into the reflecting state. Because the reflecting-state transition is well-resolved from the transmitting state transition, the pumping beam is too far off-resonance to excite the atom, and so the atom will remain in the reflecting state.

Because the pumping beam is on resonance, it does not need a high intensity to pump the atoms in the transmitting state. Once the atom is in the reflecting state, the pumping beam is now a weak, far-off-resonant beam, which has little effect on the atom, either as a pumping beam or as a conservative potential. Because its contribution as a potential is so weak, we often ignore it, but we should point out that, to atoms in the reflecting state, it is blue-detuned, and so any effect it does have serves to supplement the barrier beam in repelling atoms in the reflecting state.

With this setup, shown in Figure 4.2, as long as the atoms all start in the transmitting state, they will remain in the transmitting state, and hence be able to pass through the barrier, until they pass from the left-hand side to the right-hand side. Atoms that approach the barrier from the right-hand side will be pumped to the reflecting state before crossing the barrier, and so will not cross. Atoms that approach the barrier from the left-hand side will cross and then be pumped to the reflecting state, so the selective barrier, combined with the pumping, becomes a one-way barrier.

The main catch is the idea of being detuned far enough to act as a conservative potential. According to Equation (II.9) and Equation (II.10), the ratio of the dipole potential to the scattering rate is proportional to the detuning Δ. Thus, if we can make the detuning arbitrarily large compared to the transition linewidths, we can make the scattering rate negligibly small while increasing the intensity to get the potential barrier high enough. It should therefore be possible to have a large potential barrier without scattering, *if* we can detune the beam arbitrarily. We have a rubidium 87 trap in our lab, and limited ourselves to using the D_2 line, which fixes the three levels we will use. With those levels, the level splitting is about a thousand linewidths, which we figured would be enough that we could ignore the effects of scattering. This turned out not to be the case, as we will both discuss and demonstrate in Section IV.4.

Optical One-Way Barrier Implementation

Our one-way barrier was implemented using rubidium 87. The energy levels for the D_2 line we used are shown in Figure 4.1, while the frequencies we used for the actual barrier and pumping beams are shown in Figure 4.3. Note that the barrier beam is not tuned symmetrically between the two ground-to-excited state transition frequencies. This is intentional, and the reasons are explained when we discuss the effects of off-resonant scattering in Section IV.4.

One important thing to note in Figure 4.1 is that the frequency difference between the two ground-to-excited state transitions is fixed at about 6.8 GHz.[2] In

[2] The ground-state splitting is actually known to over 13 decimal digits of accuracy, starting out

order for the one-way barrier to work as described, the barrier beam frequency is constrained to be between the two transitions described, and hence must be at most half of 6.8 GHz, or 3.4 GHz. The linewidth of the transitions is about 6.1 MHz, so that the barrier detuning from both ground-to-excited state transitions is limited to being at most about 560 linewidths. We had originally thought this to be sufficient for avoiding scattering. As described in Section IV.4, this separation is not enough for ideal operation, and we had to resort to some tricks to deal with scattering.

Since we already had a MOT and several rubidium vapor cell saturated absorption spectroscopy setups, adding the barrier and pumping beams for the one-way barrier was not difficult. The pumping beam, as shown in Figure 4.1, happens to be resonant with the $F = 1 \longleftrightarrow F' = 2$ transition. The purpose is to pump from the transmitting to reflecting states of the atom, which happen to also be the dark and trapping states in our MOT. Thus, the optimal pumping beam transition and frequency is the exact same as the repumping beam for our MOT. Thus, we already had an ideal pumping beam; all we needed to do was pick off part of the repumping beam and direct it to the barrier beams. This pickoff is shown in Figure 3.4. In that setup, there is an AOM that we use as a rapid shutter for the repump beams. We use the first-order beam from that AOM for the repump beams in the MOT, and we pick off the zero-order beam for the pumping beam in the barrier. This way we do not decrease the repumping beam power in the MOTs, and since the AOM is off (all the MOT beams are off) while we run experiments with the barrier, all the beam power is in the zero-order and hence available for the pumping beam. The first-order beam from the shuttering AOM of the repump setup is the correct

as $6.834\ldots$ GHz (it rounds to 6.835 GHz) [51, 88]. However, the *transition* frequencies depend on the excited-state splittings too. The transition frequencies from the two ground-state manifolds to a specific excited state are split by 6.835 GHz, but the spread of available transitions from either ground state to the various excited states covers at least 200 MHz, so the actual difference between transitions varies from about 6.3 GHz to 7.0 GHz. Thus, while the ground-state splitting, is well known, that only makes the transition between actual states well known. Here, the ground "states" and excited "state" are actually multi-state manifolds, and while 6.8 GHz is a decent value to quote for the difference between those frequencies, there is a decent case for saying even that has too many significant figures, and it certainly is misleading to quote more. We will continue to quote 6.8 GHz in this section, but keep in mind the relatively large effective error bars on that number.

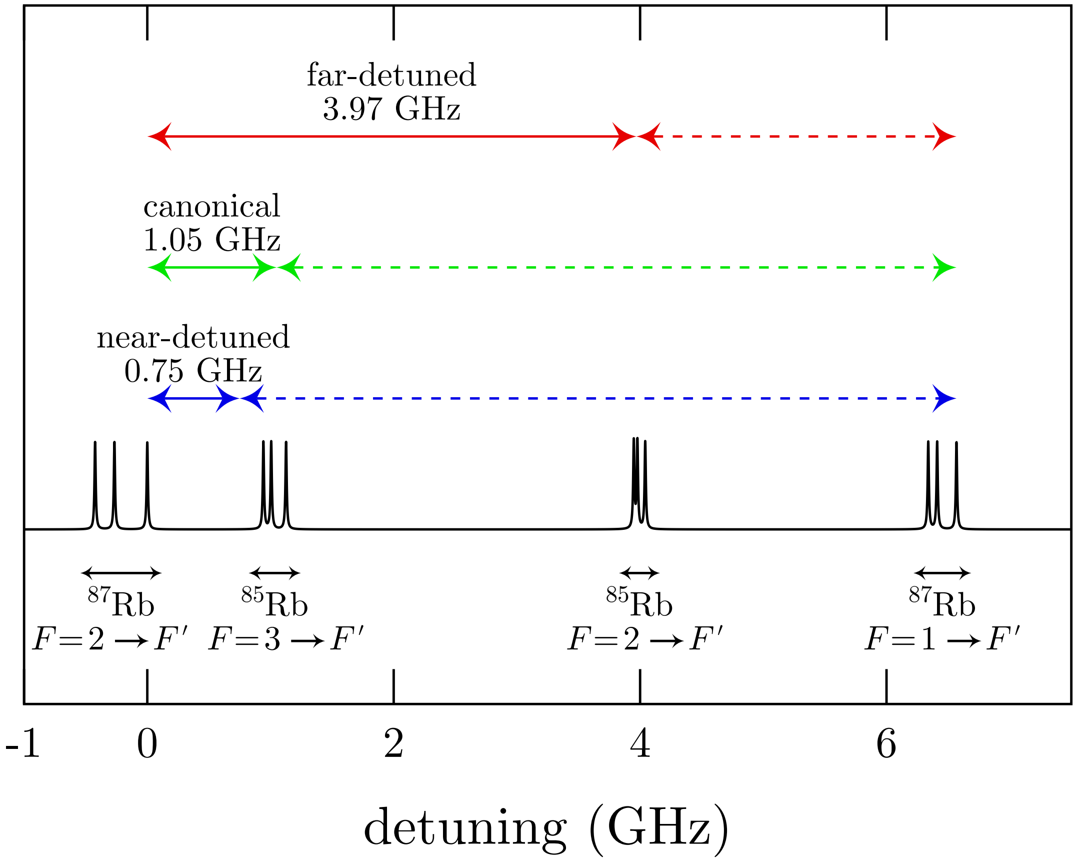


Figure 4.3. A simulated rubidium D_2 spectrum from a saturated absorption spectroscopy setup. The two ground-to-excited transitions that we use are the left-most and right-most clusters of lines, but note the two extra clusters. These are the equivalent D_2 transitions in the second common isotope of rubidium in our vapor cells, rubidium 85. Marked on the spectrum is the ideal, approximately midway barrier transition point (the "far-detuned" marked). Due to the convenient proximity of this point to the $F = 2$ rubidium 85 transitions, we used this rubidium 85 transition as our initial barrier frequency (the *middle detuning*). When we found that did not work, and figured out why, we decided we needed the barrier frequency to be closer to the reflecting-state transition frequency, and the *other* ($F = 3$) rubidium 85 transitions were helpfully at about the right frequency, and we ended up using the $F = 3 \longleftrightarrow F' = 3, 4$ crossover peak as our barrier frequency for the majority of our experiment (the *typical detuning*, marked "canonical"). When testing various detunings, we also tried a detuning that was closer to the reflecting-state transition (the *near detuning*).

frequency, which means the zero-order is not, but this allowed us to add a second shuttering AOM for the pumping barrier beam. This AOM is the same model of AOM (IntraAction) as the shuttering AOM in the repump beam setup. We use the first-order beam from that second AOM, which is the correct frequency, which allows us to control the intensity of the beam and rapidly shutter the pumping barrier beam without having to toggle the repump beams (which would not work well since the zero-order is not even close to completely distinguished when the AOM is on).

The barrier beam required its own master laser, but with a fairly minimal setup. The laser started with the standard optical setup shown in Figure 3.1, with a saturated-absorption-spectroscopy setup as shown in Figure 3.2. As with the repump laser, we dithered the laser current, and so did not have the double-passed AOM shown in the saturated absorption spectroscopy schematic. The beam, which we usually locked to one of the rubidium 85 transitions, was then coupled into a fiber with a one-to-one telescope. The fiber guided the laser beam over to the vacuum chamber, where we had another optical setup (shown in Figure 4.4) to focus the beams into the vacuum chamber. We use a mechanical speaker shutter (described in Section III.9) in the one-to-one telescope before the fiber to turn the beam on and off. This turned out to be sufficiently reliable that we could turn the beams off and image the atoms in the chamber without catching the barrier beams illuminating the vacuum chamber cell walls, without needing to wait so long that the atoms moved.

The barrier and pumping beams were carried by fiber to a small breadboard table with an optical setup shown in Figure 4.4 that combined and focused the beams into sheets. The two beams were combined on a 50/50 beamsplitter, with one output port blocked by a beam block.[3] The beams exiting the other output port were passed through an anamorphic prism pair that expanded the beams by

[3] Originally, we started with a polarizing beam splitter, and had the barrier and pumping beams linearly polarized but orthogonal to one another, so that the two beams would combine and come out mostly one port of the polarizing beam splitter. We switched to a 50/50 beamsplitter while we were still working out the details of getting the barrier to work well, so that we would not be restricted to having the polarization fixed by the polarizing beam splitter.

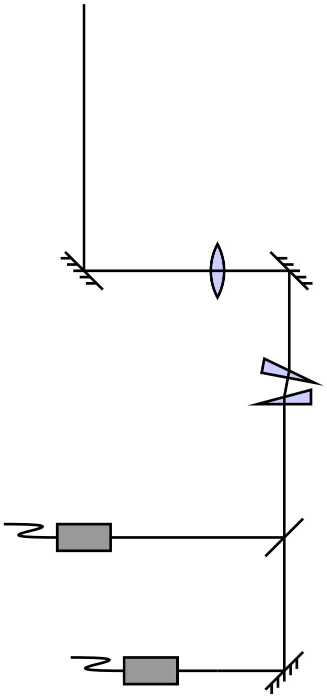


Figure 4.4. A schematic of the barrier-table optics that focused and aligned the barrier and pumping beams. The two beams are brought over by single-mode fiber, and are combined on a 50/50 beamsplitter (the extra beam is blocked). These two beams pass through an anamorphic prism pair to horizontally spread the beams by a factor of 6, so that the lens would focus them tighter in the horizontal direction than the vertical direction. After the prism pair, the beams pass through a lens and then a final steering mirror which allowed us to aim the beams directly at the focus of the dipole trap in the chamber. The barrier beam was reflected off of one mirror that allowed us to change the angle of the barrier beam by small amounts, which translates to a separation of the foci of the beams after the lens. The beams are not normal to the cell because we also wanted the absorption imaging beam to be able to image the atoms without having the barrier optics getting in the way, and without the barrier beams shining into the camera lens.

a factor of 6 in the horizontal direction. This expanded beam was reflected off a mirror, passed through a 250 mm focal length planoconvex lens, and then off a final mirror. The final mirror gave us steering capabilities to makes sure two beams crossed through the focus of the dipole trap. The second-to-last mirror simply helped keep everything on the table.

The barrier beam used a steering mirror to before the beamsplitter to let us fine-tune the angle of the beam separately from the pumping beam. Beams passing through a lens focus to a location that depends more on the angle of the beam than the physical offset of the beam from the axis of the lens. Having slightly different angles for the two beams let us adjust the separation of foci of the barrier and pumping beams separately. While we did develop ways of measuring the separation of the two beams with surprisingly good accuracy, which we will describe later, the methods took some time. Because of this slow-ish measurement, we performed a rough calibration of the steering mirror, so we would know about how far to rotate the adjustment knob on the mirror mount to change the separation by a desired amount. Thus, by measuring the separation once, we could then adjust the steering mirror by an appropriate amount, and get quite close to our desired separation, usually with only two measurements, once before and once after the change.

The rotations required for the separation adjustments were quite small, sometimes on the order of 1°. To facilitate such small rotations, we devised a simple system. As shown in Figure 4.5, the adjustment knob on the steering mirror has small grooves, presumably for grip. There are about 80 grooves around the knob, so each groove represents about 4° of rotation. We mounted a very sharp pin on the mirror mount, such that it pointed at these grooves. Using a thin stick and the ink from some pens we had in the lab, we painted one groove red, left the grooves to either side blank, and then painted the next groove to either side blue. By inspection, we could count how many grooves were between the needle and the red groove, and thus get the position of the knob to about 4°. Using a jeweler's glass, we could determine where the needle was within the groove to approximately 1/4 groove, for an accuracy of about 1°. Our rule of thumb was one groove (about 4°) resulted in an 8 μm change in separation between the barrier beam and the pumping beam at

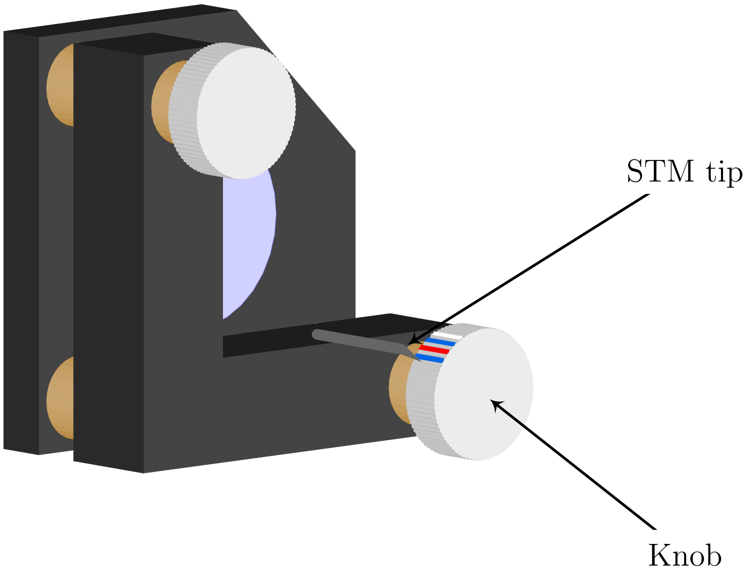


Figure 4.5. A schematic of the horizontal adjustment knob for the barrier beam steering mirror on the barrier table. The knob has grooves in it to provide a better grip, which we used to estimate beam separation. By marking several of the grooves with colors, and using a very sharp needle as a pointer, we were able to measure knob rotations of about 1/4 of the turn from one groove to the next, or about 1/320 of a rotation.

the barrier beam focus. With the jeweler's glass, knowing the current beam separation, we could then change the separation to any value we used to within about 2 μm, which was almost our measurement accuracy, and close enough for most of the separations we wished to attain.

We had to deal with some difficulties for the barrier and pumping beams. To reduce scattering, and to use beam power most efficiently, we wanted the barrier beam as narrow as possible where it crossed the focus of the dipole trap. We wanted the pumping beam as close to the barrier as possible, but without extending through the barrier, and without much, if any, overlap, as changing the state while the beams were crossing the barrier could be a problem. One way to do this is to focus that beam to a very narrow sheet as well. We also wanted the two beams very close to each other. If they were far apart, some atoms could oscillate around the focus of the dipole trap with a small amplitude, and never reach the pumping beam. As it turns out, for reasons we will discuss in Section IV.4, we actually wanted a tiny amount of overlap between the two beams. The simple reason is the barrier beam had a slight tendency to pump atoms to the transmitting state while they reflected off of it, ruining the barrier. A small amount of overlap with the pumping beam countered this, at the expense of a little extra heating.

The net result is we wanted the beams to be on the order of 10 μm wide along the axis of the dipole trap, more than 30 μm tall to cover the whole focus of the dipole trap, and separated by about 30 μm. Obviously, the optics need to be outside the cell of our vacuum chamber, which is about 30 mm wide, so we need to be at least 15 mm away from the atoms. In order to get a tighter focus, and to keep the optics far enough away that they did not block either the MOT beams (for fluorescence imaging) or the absorption imaging beam the illumination beams (for trapping and imaging), we used a 250 mm focal-length lens, so the optics needed to be about 250 mm away from the dipole-trap focus, where the atoms would be.. From Gaussian beam optics, we can compute the Rayleigh length of the beams with this focus (and wavelength 780 nm), with the result that, in order to get a good focus, the lens needs to be placed at exactly the focal length away from the focus of the dipole trap, plus or minus about a quarter of a millimeter. The target area is the focus of the dipole trap, which, as described in Section III.7, is about 30.9 μm high (plus or minus the height of the focused barrier beam and pumping beam, which gives us a little more leeway). Along the axis of the beam, we want to be in the center, accurate to much less than the Rayleigh length of about 2.8 mm, say to within about a third of a millimeter. For comparison, imagine we are shining the lasers the long way across a football field instead of the 250 mm focal length. In that case, we are shining two beams onto a target that is about 10 cm by 1 cm, 100 m away. That is about the size of a candy bar, across the long dimension of a football field. Not only that, but the beams need to each be focused down to a width of about half a centimeter, and separated by about a centimeter. These are impressive, but since we are building on something as solid as an optical table, that kind of stability and precision is not all that difficult.

The difficulty is that we need to hit the candy bar, get our distance to the candy bar to within a fraction of a centimeter, and get the beams about a centimeter apart on the candy bar, *without being able to see the candy bar*. The focus of the dipole trap is inside the vacuum chamber, and we cannot place any optics in there to mark it, and obviously the barrier and pumping beams do not interact with the focus of the dipole trap.[4] We used two separate tricks to align the barrier and pumping beams. The first trick was to check the width of the beams at their focus, and their separation at their focus, outside the chamber. Once the focus and separation is to our satisfaction, we can the use the final mirror to hit our target. The second trick was to tell where the target was by trapping atoms in the dipole trap, because the atoms, unlike the dipole trap itself, *do* interact with the barrier and pumping beams. Unfortunately, they cannot be contained to a very small region, so it is like trying to hit the candy bar across the football field, but with blurry vision.

To do this, we needed a way to measure the beams away from the vacuum chamber, in order to check their widths at the focus and set their separation. Once in place, we would also need to change the distance of the barrier optics, so that the focus of the barrier and pumping beams would be where the two beams crossed the dipole trap focus. By placing all the barrier beam optics on a separate, portable table that we could bolt on our optics table, we allowed ourselves a way to move the barrier optics away from the chamber and check the beam widths and separation accurately right at their focus. The light for the two beams came out of fiber couplers, so we could move the table without needing to realign any beams.

Initially, we would bolt the table down somewhere with few obstructions on our optical table, and place a borrowed beam profiler near the focus of the barrier and pumping beams. We would then use that to measure the beam waists at the focus, and the beam separations. Our initial table layout was simpler than that shown in Figure 4.4, but we quickly found out that the beams were too wide, with $1/e^2$ intensity radii on the order of 100 μm. Later on, we found that our adjustable, movable fiber mounts tended to be susceptible to small bumps, which would shift

[4] The vacuum chamber is very heavy, has a lot of parts, has some optics and coils around it that get in the way, and even a small bump of the Hellma cell could break it, setting us back months, so moving the chamber was not an option.

them just enough to change the beam separation significantly. To combat these two problems, we redesigned the barrier table to what is shown in Figure 4.4. The fiber couplers were mounted solidly to the table, and did not have adjustable mounts. We found that we could get the two beams close enough to parallel by moving the mounts around by hand, and then clamping them to the barrier table. We would then adjust the separation with the one mirror, and double-check that the beams were still close enough to parallel. Since we needed to be within less than a millimeter of the focus of the beams, we only needed them to be parallel enough that the separation did not change much over a millimeter or two, and a few iterations of making the beams parallel, clamping the mounts to the barrier table, setting the separation to around 20 μm to 30 μm, and then making sure the beams did not cross within about a half inch of the focus was usually sufficient. To get a tighter focus, we added an anamorphic prism pair to the table. This expands the beams horizontally by about a factor of 6 before they enter the lens. Even when not diffraction limited, lenses have a tendency to follow the same rule as the diffraction limit, where the larger your iris size, the smaller the spot you can resolve. Having a wider beam enter the lens helped it to focus tighter. That, plus using larger optics to prevent clipping of the beams and using better collimators so that the beams were closer to collimated, allowed us to get the beam waist $1/e^2$ radii down to $11.5\,(5)$ μm for the barrier beam and $13\,(2)$ μm for the pumping beam. However, since the prism pair only expanded the beam horizontally, the beams did not focus nearly as well vertically, with vertical $1/e^2$ radii of $80\,(7)$ μm for the barrier beam and $60\,(7)$ μm for the pumping beam.

The anamorphic prism pair was custom-built by mounting a pair of Edmund Optics 30°–60°–90° uncoated prisms (NT43–648 and NT43–649). We calculated the appropriate mounting angles for the prisms, and machined some grooves on a block of aluminum at those angles. We then placed the prisms in those grooves, pressed up against the edges, epoxied them in place, and mounted the aluminum block on a standard optics mount.

Measuring the beam waists and separation evolved as we got the beam waists narrower and tried to get the beam separation as stable as possible. At first, the beams were large enough that we could measure their waists directly with a borrowed beam profiler. We would move the barrier table away from the vacuum chamber, so that the region where the beam foci were was accessible, and place the beam profiler sensor there. We mounted the beam profiler sensor on a micrometer stage, so we could move it known amounts away from and towards the barrier table, enabling us to locate the focus, and measure the change in beam separation as a function of distance. The software that came with the beam profiler quoted beam waists and positions, which we recorded by hand, giving us waist and separation measurements. The micrometer position that gave the smallest waists was where the foci were. However, the small waist sizes we achieved with the anamorphic prism pair were on the order of the pixel size of the beam profiler, so we needed a more accurate measurement.

After we could no longer reliably use the beam profiler, we settled on an advanced version of the razor method of measuring beam waists. This version entails sweeping a razor blade edge across the beam, with a power meter measuring the amount of the beam that passes the edge. Because a narrow beam diffracts quite a bit off a razor's edge, the power meter needs to be quite close to the razor. The simplest version of this method is to assume a Gaussian profile, and measure the razor positions where the razor blocks 10% and 90% of the light (or 20% and 80%, or some other value). As the razor sweeps across a Gaussian profile, the amount of light blocked is the integral of the Gaussian profile, or an error function. Using this, we can compute a scaling factor that, when multiplied by the distance between the 10% and 90% points, gives the $1/e^2$ intensity radius. For the 10% and 90% measurements, multiplying the separation by 0.780 gives the $1/e^2$ intensity radius (1.188 for 20% and 80%). This method can be improved upon by taking a few extra points. Indeed, with enough points, a one-dimensional integrated intensity profile can be measured.

In our case, we placed the razor blade on our computer-controlled air cart, allowing us to automate taking many points to recreate a one-dimensional beam profile. The blade was also mounted on a micrometer, so we could shift towards and away from the barrier table, allowing us to find the focus of the beams, and measure how rapidly the separation changed as we moved away from the foci. In fact, we found that the air cart was capable of traveling smoothly at a fixed velocity, so by using an oscilloscope to measure the output of the power meter as the air cart swept the razor through the beam, we could get a one-dimensional beam profile in one quick sweep. We could then get the beam waist by fitting a curve to that profile. Repeating the measurements with different sweep speeds, and comparing with the method where we stopped the cart at each point, we found this method repeatably gave the same beam waist down to the sub-micron level.[5] Since the air cart is supposedly accurate down to that level as well, we believe that is an absolute accuracy. The only thing we needed to watch was that the sweep was slow enough that the power meter could keep up with the rapidly changing power (the smaller the beam, the faster the razor cuts the power off). We could not get beam positions with this method, as we did not know exactly where the cart was at a given time as measured by the oscilloscope, so instead we implemented a binary search algorithm with the air cart. We would measure the beam power with the beam fully blocked and fully un-blocked, and then bisect that interval in half, until we found the point where the exactly half the power was blocked by the razor. This position is relative to an arbitrary reference point, but as long as we found the centers of both the barrier beam and the pumping beam without moving that reference, we could accurately find the center Our estimated error comes from repeating the same measurements with different intervals for the bisections, and seems to be well under a micron (for small beam waists).

At first, we thought this method was working well. We would carefully measure the beam waists, and adjust the separation to something reasonable, and then move the barrier table back near the vacuum chamber, and align the beams on the atoms, a tedious process which we will describe shortly. We would then see if the barrier worked, and typically, it did not. We would then pull the table back away from the chamber, and re-measure the waists and beam separations in preparation to set the

[5] This is how we measure horizontal waists. Because the air cart only sweeps one direction, and we did not want to rotate our setup, we used a beam profiler to measure the vertical waists. The beam profiler did not have sufficient resolution to measure the horizontal waists, but worked well for the vertical waists.

separations to a different value. The beam waists were consistently the same, to within a small error, but the separation was almost always different.

Initially, we thought we had bumped something to misalign the beams when moving the table. We checked that all the optics were solidly mounted and resistant to gentle nudges. The schematic in Figure 4.4 shows the final revision; prior versions with different layouts also used less-study mounts. Eventually, after several moves, we ran out of things that could be the problem, and decided that we needed a way to measure the beam separations with the table *in situ*. While preparing this method, we noticed that the very act of clamping and unclamping the barrier table to the optical table flexed it enough to shift the beam separations. This meant that we could not pull the table away from the vacuum chamber to measure beam waists and beam separation.

We knew from all the times we had moved the table that the beam waists were quite stable, so we really just needed a way to check the beam separations, and we settled on a cheap beam profiler, described in Section III.9. We basically had a cheap webcam called the PetCam where we removed the lens and added strong neutral density filters to it. This let us shine beams directly on the sensor of the camera, imaging the beam directly. Sensors from cameras such as this are typically small, with pixels of about 5.6 μm on a side, which makes them surprisingly good at profiling rather small laser beams. The camera itself is rather noisy, and images taken from it use a rather lossy JPEG compression, but by subtracting an averaged background and averaging multiple images of the beam, we can produce a decent-quality image, certainly good enough to find the beam size and center. The powers of the barrier and pumping beams were also quite different when operating the one-way barrier; we changed some ND filters at the input of the barrier beam fiber and used the AOM that adjusts the pumping beam power to roughly match the two beam powers, and to help them image nicely on the camera.

We mounted the camera on the micrometer translation stage with a magnetic mount, so we could quickly and repeatedly insert and remove the camera. By checking how the beams appeared to move on the camera as we translated the camera in various directions, we were able to verify that the pixel size was indeed

5.6 μm on a side, and that we had the camera mounted pretty close to parallel to the table, so that our separation measurements would be a good approximation to the actual separation seen along the dipole-trap axis. We could insert a small mirror between the barrier table and the Hellma cell, and reflect the beams out approximately parallel to the Hellma cell. The camera was small enough to fit in there approximately where these reflected beams focused, allowing us to measure the beam separation as though the camera were at the focus of the dipole trap.

This camera allowed us to do both fine and rough alignments. By watching the near-real-time webcam feed, we could easily find the approximate location of the beam foci by scanning the camera along the beam axis and looking for the narrowest beams. As shown in Figure 4.6, we could easily see the two beams with their asymmetric profile, allowing us to quickly align them vertically to each other, and get the horizontal separation about right. We could then do a more refined measurement. We would check one beam at a time (by physically blocking either the barrier or the pumping beam), and take images of the beam as we scanned the camera through the focus of the beams. An automated script would then crop the images, do column and row sums to create one-dimensional profiles, and curve-fit the results, giving beam centers and waists as a function of distance from the barrier table (up to some unknown overall constant). We could then plot the horizontal positions and beam waists of both beams together on the same plots, as functions of camera location. One such data set is shown in Figure 4.7, where we can see that the camera could not reliably resolve the horizontal beam waist very well near the foci of the beams; the camera could resolve the vertical beam waists well enough that we used the camera for our quoted vertical waist measurements. Gaussian optics tells us that the waist as a function of distance from the focus should be hyperbolic, which means the waists are asymptotically linear with respect to distance, and those linear curves intersect at the focus, with a waist of zero. By fitting lines to the outer regions of the waist versus position curves, and looking at their intersections, we can infer, reasonably accurately, where the focus of the beams are. Then, by looking at the beam centers versus camera position, we can read off the beam separation at the focus, along with how rapidly that separation changes with camera position.

Naturally, before installing this system near the vacuum chamber, we tested it out in the open, where we could compare results with the razor mounted on the air cart, and the agreement was typically within a micron, probably limited to the resolution of the camera. This agreement was good enough for our purposes.

Interestingly enough, the camera did have enough resolution to measure the vertical waists of the beam, which were much larger. Thus, we could actually check the vertical size of the beams, which we could not do an air cart that could only move horizontally.

With this camera, we could measure (and set) the beam separation at the foci of the barrier and pumping beams, and every so often, we would pull the table out and double-check that the beam waists had not changed. Then, when we placed the table back, we would need to align the beams to the dipole trap and set and check the separation again.

We will now outline the alignment procedure, which was rather tedious. We clamped a bar to the optical table parallel to the approximate path what we wanted the barrier and pumping beam to take. This bar was placed so that we could press the barrier table legs up against it. By sliding the barrier table along this guide, we could move the focus of the barrier and pumping beams in and out of the Hellma cell. We could then control the direction of the beams with the last mirror on the barrier table, and the distance from the focus of the dipole trap by sliding the table. By comparing where the beams entered and exited the Hellma cell, we could get a rough alignment. We would then turn on the MOT, and center it on the focus of the dipole trap, as described in Section II.3. With the dipole beam off, we would then tune the barrier beam to the rubidium 87 MOT transition, and sweep the frequency through the transition. Because the barrier beam, at resonance, strongly affects the MOT, we did not need a high intensity, so, as with aligning the dipole trap, we could move the table back far enough that the barrier beam would be out of focus where it crossed the dipole trap. This gave us a reasonably large beam to try and hit the MOT. Actually hitting the MOT was then not too difficult, and we could easily see the MOT flicker (as the barrier beam swept through resonance) in the security camera that was trained on the MOT.

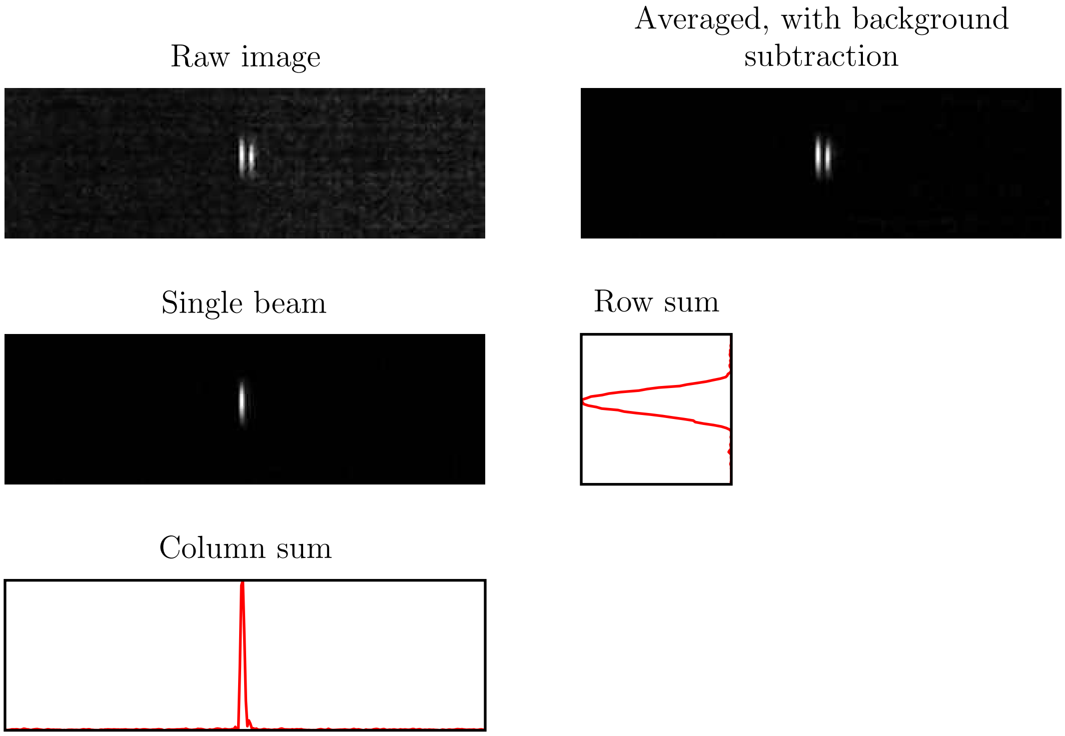


Figure 4.6. Viewing beam profiles with the PetCam. The raw image shows the two beams as they would be seen viewing the PetCam images with a web browser. The image quality is greatly improved after averaging multiple images together and subtracting the background. Ordinarily we viewed both beams when performing coarse alignment, which allowed us to align them vertically and set the horizontal separation to the rough value we might want. When taking data for measuring beam separation, we always blocked one beam so we could measure one beam at a time. We did not save or average images for the coarse alignments, so most of our data images show only one beam. Here, we have combined images of the two beams that were taken separately to simulate the sort of image we would get if neither beam were blocked. The asymmetry of the beams is a result of the anamorphic prism pair shown in Figure 4.4. The wider axis of the beam focuses to a narrower waist, creating the vertical sheets of light we wanted to cross the dipole trap. Below the images, we show a single beam (where we have physically blocked the other beam), and the corresponding one-dimensional profiles, summed over the region where the beam is. We fit Gaussians to these profiles to determine the beam waists (vertical only) and center location (horizontal and vertical).

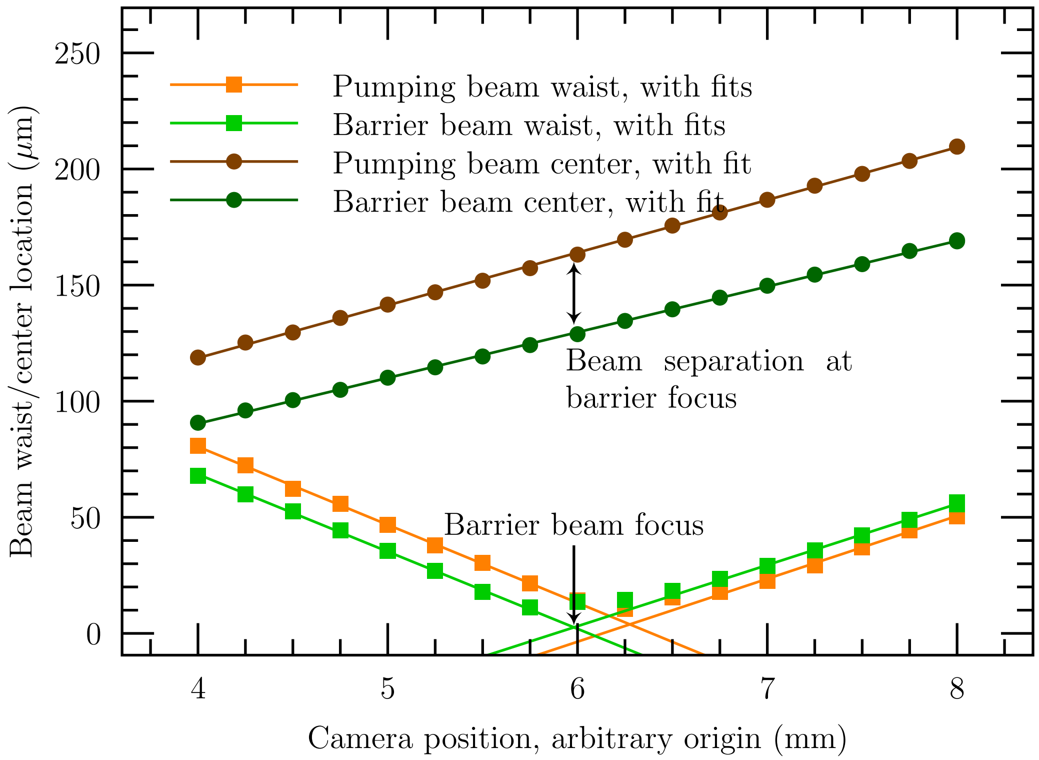


Figure 4.7. Finding the beam separation from the PetCam profiler. Using a series of curve fits, such as is shown in Figure 4.6, we can plot beam waist and beam position (on the camera sensor) as a function of camera position along the beam axis. Here, we show some data from both the barrier and pumping beams plotted on the same axes. The beam waists near the foci are actually between 11 μm and 14 μm as measured by the razor mounted on the air cart, but the camera cannot resolve such a narrow beam due to its limited resolution. Because the vertical waists are larger, we were able to use similar data for the vertical waists as measurements of that length. However, by looking at the asymptotic parts of the waist measurements, we can infer the location of the focus, and measure the beam separation there. Since it is the barrier beam that really needs to be narrow, we try to align to that focus in the experiment, so we measure the separation at that point here. We note that the separation is not much different at the pumping beam focus. The slopes of the waist fits are not symmetric about the focus, which we assume is related to the beams not being perfectly Gaussian.

We would then stop the ramping the frequency of the barrier beam, and set it to a little off of resonance, and begin iterating. We would first sweep the beam directly through the MOT, and see how much the atoms were displaced by the beam. Then we would shift the table one way or the other, and see if we could affect the MOT more. As we got the focus of the barrier beam closer to the MOT each time, the beam would be more intense, and hence affect the MOT more. When it completely destroyed the MOT, we would tune the beam a little further from resonance to reduce its effects, and continue with the process.

Eventually, we would reach the point where the beam was so well focused at the MOT that it would not affect the MOT very much. We would try to find this point on each side of the focus, and read the table position off a rule clamped on the table guide. After that, we would place the table at the midpoint of the two positions, and begin centering on the focus of the dipole trap itself.

This part was more difficult, because we cannot image atoms in the dipole trap without disrupting them (and the dipole trap tunes them out of resonance anyways, making imaging more difficult). The basic process is we would tune the barrier beam to the rubidium 87 repump line, test it's effectiveness as a barrier. At this frequency, any atoms in the lower ground state are pumped by this frequency to the upper ground state, just like how the repump beam is used in the MOT. Once in the upper ground state, the beam acts as a roughly conservative, repelling potential. This is not a one-way barrier, but a barrier nonetheless.

The general procedure was to load the MOT into one side of the dipole trap and release them with the barrier beam on and crossing the dipole trap. After roughly a quarter to a half a period, we would turn the dipole trap and barrier beams off and image the atoms. If the barrier beam were not crossing the dipole trap, the atoms would be seen on the far side of the dipole trap. If the barrier beam were crossing the dipole trap, at least some of the atoms would have reflected and be seen on the same side of the dipole trap as we had loaded the atoms. The limiting factor in our alignment was our ability to place the table and point the beams, so we did not need a really accurate measurement of how many atoms reflected. As such, since we needed many iterations for this alignment, we just repeated the experiment twice for each of two times between about a quarter of a period and a half of a period (and again for dark frames to subtract images). Some sample images are shown in Figure 4.8, where we can clearly see the barrier. Looking at the images, we could first adjust the beam location horizontally, to try and get it roughly in the center of the dipole-trap location, repeating the imaging after each adjustment. We would then adjust vertically, repeating the imaging after each adjustment, looking for the location that gave us the best reflection. As an example of the sort of comparison we needed to make, the two sets shown in Figure 4.8 shows what we could see with a near-optimum position, versus one where the barrier beam mirror was tweaked just about the smallest amount we could, twice.

Note that not all the atoms reflect off of the barrier in Figure 4.8. This is because the kinetic energy of the atoms is comparable to the height of the barrier. The more energetic atoms can cross the barrier; the less energetic atoms reflect.[6] The higher the intensity of the barrier beam, the higher the potential (which is proportional to intensity) it presents, and the larger fraction of the atoms it repels. The barrier beam, near the focus, is roughly a vertical sheet (as seen in the beam-profile shown in Figure 4.6). If the center beam passes above or below the dipole trap, the maximum intensity of the barrier beam seen by atoms in the dipole trap is lower than if the barrier beam passes right through the center of the dipole trap, so comparing how many atoms reflect indicates which beam location is closer to optimum. Likewise, if the beam is more out of focus, the power is more spread out, and the maximum intensity is lower, so fewer atoms reflect.

Once we got the best reflection possible, passing through the region once while moving the barrier beam one direction, and returning to it going the other, we would shift the barrier table a small amount either towards or away from the Hellma cell, and repeat. If we could find a better reflection at that table position, we would

[6] Actually, the motion of the atoms perpendicular to the trap axis also matters. As described in Section IV.5, when we tried to simulate the motion of the atoms in the trap, we found out that atoms rapidly orbiting the dipole-trap axis see a different effective potential, which helps to repel them from the focus. Thus, atoms with higher angular momentum are more likely to reflect than atoms with lower angular momentum, even if their initial speed along the dipole-trap axis is the same.

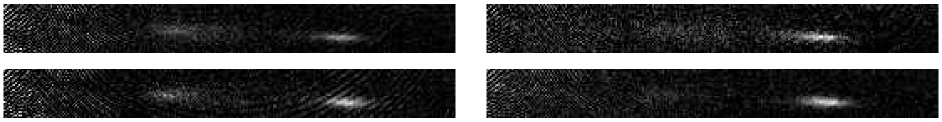


Figure 4.8. Sample images used for aligning the barrier beams vertically onto the focus of the dipole trap. The left column shows two images, 14 ms and 16 ms after releasing the MOT atoms in the dipole trap, where the barrier beam, tuned to the rubidium 87 transition, is near optimally aligned. To help with alignment, the beam is attenuated, as a higher power beam would reflect everything whether the alignment was optimum or not. The right column shows a similar set, where the barrier beam was not as well aligned. We can clearly see the atoms just after the process of passing through the barrier, with some being reflected. The better-aligned the barrier is, the more atoms get reflected. Each of these images is the average of two repetitions, with frames without atoms subtracted.

move it further in that direction; otherwise, we would change directions. We would repeat process until we passed through an optimum. Sometimes, we would reach a point where all the atoms were reflected. When this happens, we would add some neutral density filters to attenuate the barrier beam. Once all the atoms reflect, we cannot detect a better alignment, so it was in our best interest to keep some, but not all, atoms reflecting. If we were just starting, or if we got lost, we could always increase the power, so that we could get some reflection even if we were not closely aligned. For our dipole trap and barrier beam size, we found that a barrier beam power of a little under 1 mW would reflect most of the atoms when aligned as best as we could get it.

As previously mentioned, the Rayleigh length of the barrier beam and pumping beam is on the order of a quarter of a millimeter, so we want to get the table in the optimum location, to within a quarter of a millimeter. Typically, we would move the barrier table in increments of half a millimeter, and once we found the best location, we would clamp it there. If we found two locations where we could not determine which was better, we would try to clamp the table at the quarter millimeter increment between the two locations, giving us roughly quarter-millimeter precision.

After aligning vertically and longitudinally, we still needed to align horizontally.

This test was relatively simple. We would load the dipole trap for a while and let the atoms oscillate a little before imaging. We would repeat this measurement both with and without the barrier beam on, and compare the images, as shown in Figure 4.9. The barrier beam would produce a sharp cut in the profile of the atoms, which we would make sure was near the center of the profile of atoms in the full trap. We were careful to load the trap and let the atoms sit long enough that the atomic distribution had settled into something symmetric about the dipole trap focus, so the center of the distribution should correspond to the dipole-trap focus. If the alignment was close enough, we would take it. If not, we would use the mirror on the barrier table to try and center the beam, repeating this measurement with each adjustment. Once we were done with the adjustment, we would go back and check the vertical alignment. The horizontal, vertical, and longitudinal adjustments were only weakly coupled to each other, and adjusting one rarely required much change in any other.

This alignment procedure was tedious, but, once complete, checking for optimum alignment was not too terrible, as we were already near the optimum location. Even when we did pull the table out, if we put some stops up against the table legs before moving it, we could usually place it back close enough that re-aligning it was not too difficult.

We finish this section by mentioning that for the vast majority of the data we took once we got the barrier working, the main barrier beam was linearly polarized parallel to the axis of the dipole-trap beam, and the pumping barrier beam was linearly polarized perpendicular to the axis of the dipole-trap beam.[7] We started with these polarizations because we were originally combining the two barrier beams with a polarizing beam splitter. As seen in the barrier table schematic (Figure 4.4), the main barrier beam needed to transmit through the beam splitter while the

[7] Note that these polarizations are reported *incorrectly* in one of our publications [20]. That paper reports that the main barrier beam was normally perpendicular to the dipole-trap axis (it was parallel, or horizontal), and the pumping barrier beam was normally parallel to the dipole-trap axis (it was perpendicular, or vertical). We also mention that we briefly tried having both beams polarized parallel to the dipole-trap axis, when we actually tried having both beams perpendicular to the dipole-trap axis.

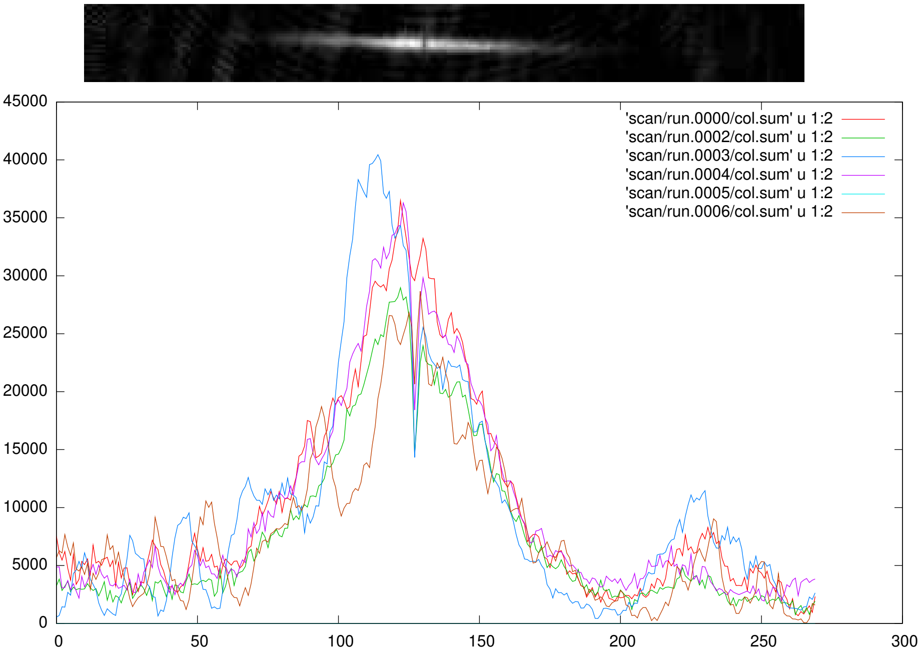


Figure 4.9. Sample images used for aligning the barrier beams horizontally onto the focus of the dipole trap. These images show a dipole trap loaded with atoms, where the loading time and relaxation time were long enough to let the atoms equilibrate. In one image, the barrier beam was off, and in the other, it was on. By integrating the images vertically, we could produce the profiles shown in the graph. The barrier clearly shows up as a dip in one graph, and we would make sure that dip occurred in the center of the profile from the full dipole trap. These images are the result of 5 repetitions, with dark frames subtracted.

pumping barrier beam needed to reflect. With such a setup, a horizontally polarized beam would transmit better, while a vertically polarized beam would reflect better; this is why we made the main barrier beam horizontally polarized and the pumping barrier beam vertically polarized. When we replaced the polarizing beam splitter with the 50/50 beam splitter that we used to take the data shown in this thesis, we recalibrated the beam powers, but left the polarizations as they were.

Experimental Demonstration of an Optical One-Way Barrier

We will now discuss the actual settings used for the barrier, and show results

[18, 20]. We start with approximately 2×10^5 atoms in the MOT, cooled to around 30 μK. The dipole trap beam, running at a typical power of around 10 W, should have a power of about 9.3 (5) W inside the chamber. Focused to a $1/e^2$ intensity radius of about 30.9 (5) μm waist, at the quoted wavelength of 1090 (5) nm, that should produce a conservative potential with a depth of $k_\mathrm{B} \times 0.9$ mK, as described in Section III.7. We load the atoms into the dipole trap, let them move around in the trap, and then later image them to see how the barrier affected them.

The one-way barrier beams cross the dipole trap at an angle of about $12\,(3)^\circ$. As described in Figure 4.4, this is just to keep the barrier beams from being in the same path as the absorption-imaging beams, both so the optics are separate and so the beams do not shine into the camera. The main barrier beam is focused to a horizontal waist of 11.5 (5) μm and a vertical waist of 80 (7) μm, $1/e^2$ intensity radii. The pumping barrier beam is focused to 13 (2) μm and 60 (7) μm, respectively. We typically had the pumping beam focused 34 (1) μm to the right (as seen from the camera) of the main barrier beam. The main barrier beam typically had a power of about 40 (4) μW inside the cell, locked to the rubidium 85 $F = 3 \longleftrightarrow F' = 3, 4$ crossover peak (shown in Figure 4.3), while the pumping beam had a typical power of 0.36 (4) μW inside the cell (locked to the rubidium 87 repumping transition). The peak intensities of the two beams are 2800 mW/cm^2 for the barrier beam and 29 mW/cm^2 for the pumping beam. In other terms, the barrier beam should represent a potential barrier of about $k_\mathrm{B} \times 0.22$ mK for atoms in the reflecting state, and a potential well of about $k_\mathrm{B} \times -0.045$ mK for atoms in the transmitting state, while the pumping beam is on the order or 10 times the saturation intensity of the rubidium 87 D_2 line.[8]

We would load the MOT for 5 s, and cool for 20 ms, as described in Section II.2. At the beginning of the cooling stage, we would shift currents in the Helmholtz coils to move the MOT from the optimum loading location to the spot where we wished

[8] We assumed the beam was resonant with the atom here, and only used an order of magnitude for the saturation intensity. In reality, the dipole-trap beam shifts the resonances of the atoms, so the pumping beam is effectively detuned by varying amounts depending on how far the atom is from the trap axis, but this saturation intensity ratio still demonstrates that the pumping beam easily saturates the atoms and should therefore pump the atoms near the maximum possible rate.

to load the dipole trap. Once the MOT is loaded and cooled, we turn on the dipole trap beam and let the MOT load into the dipole trap. For data where we wished to watch the atoms travel back and forth, we would only load for 5 ms, with the Helmholtz coils set so that the loading took place 0.95 (5) mm to one side of the dipole-trap focus. During longer loading times, atoms in the dipole beam would be Stark-shifted out of resonance with the MOT beams, and so would move about in the dipole beam without being held in place by the MOT beams. Thus, they would start oscillating back and forth about the focus of the dipole trap, smearing out the atomic distribution. For these short loads, we were unable to open the camera shutter quickly enough to catch the start of the oscillation without catching part of the MOT loading into the dipole trap, and the atoms which were not loaded drifting away. As these were unwanted additions to our data, we typically waited for the atoms to oscillate half a period (about 22 ms, not counting the 5 ms spent loading the dipole trap), and then turned the barrier beams on to start the experiment. During this half-oscillation, the atoms moved to the opposite side of the dipole trap, and dephasing due to angular momentum and anharmonicitiy reduced the amplitude of the oscillation. As a result, even though we typically pulled the atoms back 0.95 (5) mm, the initial position of the atoms in our first data frame appears to be 0.58 (8) mm on the opposite side.

We would also use this extra time while the MOT atoms fell away to pump the atoms into the correct state, either by leaving the MOT trapping beams (or MOT repump beams) on a little longer than the other MOT beam, or by pulsing one of the two sets of beams as the atoms passed through the center of the dipole trap. We verified that we could successfully pump the atoms into either state in this manner by taking absorption images of the atoms with and without the repumping beams. If all the atoms were in the transmitting state, which is not resonant with the absorption beam, then the image without the repumping beams (which pumps atoms in the resonant reflecting state) would show no atoms. If all the atoms were in the reflecting state, then the with-repump and without-repump beams should look the same.

In cases where we either wanted the atoms evenly distributed about the dipole-trap focus, or where it did not matter if the atoms distributed themselves around, we loaded for 110 ms, which loaded more atoms. Often, we loaded the same distance off to one side of the focus as if we were going to watch the atoms traverse from side to side, as that resulted in a wider distribution of atoms. We would also wait longer before turning the barrier beams on or taking images, giving more time for the atoms to settle into a relatively static distribution. To distinguish these cases from the cases where we use a short (5 ms) loading time to have a local initial distribution, we usually refer to this longer 110 ms load with a longer waiting period a *full load.*

Our raw data from an experiment is in the form of images. Figure 4.10 shows how we process raw images into something we can either publish or use as a measurement. We start with images of the absorption beam, which we tweaked to try and get it to uniformly cover the field of view.

Each image with atoms is immediately followed by a repeat of the experimental procedure, but without the MOT coils on, so there are no atoms.[9] As explained in Section II.4, the atoms scatter light out of the absorption beam, creating a shadow. By subtracting the two images, we remove the main beam, and show just that shadow. Each atom scatters a fixed proportion of the illuminated light, but if one atom scatters some light, there is less for the next atom to scatter. This causes the amount of light scattered, and therefore the depth of the shadow, to decay exponentially with the density of atoms. Technically, the number of atoms is proportional to the logarithm of the fraction of light scattered, ln(with atoms / without atoms). As can be seen by the raw images in Figure 4.10, the shadow cast is a small fraction of the light. This allows us to make a linear approximation. We computed the

[9] We found that, with our camera, the subtraction did not work well if we took a second image immediately afterwards. The time between images needed to be similar in order for a few camera artifacts to cancel out. As long as we needed some time between images, we figured we might as well duplicate the experiment as closely as possible. That way, any oddities in the illumination (such as the camera shutter opening early enough to catch a reflection from a beam that was off during the illumination pulse) would be the same, and thus would subtract. To speed things up a little, we reduced the MOT load times to 2 s, too, which we found did not alter the images.

atomic density as proportional to:

$$\ln\left(\frac{\text{total light} - \text{light scattered by atoms}}{\text{total light}}\right) = \ln\left(1 - \frac{\text{light scattered by atoms}}{\text{total light}}\right) \approx -\frac{\text{light scattered by atoms}}{\text{total light}}. \quad \text{(IV.1)}$$

This let us simply take the difference between the two images, normalized to the amount of light available.

To reduce noise, we would then repeat this pair of procedures many times and average the results together. For simple tests, we would only average two or so sets of images, while for publication-quality data, we would sometimes average over 50 sets of images; the exact amount was chosen based on how clean the images were. Noise in the images resulted from several causes. Sometimes the camera shutter stuck, resulting in either no light in an image at all, or, if it merely delayed the shutter, causes part of the image to be blocked, or large fringes to cover the image. Sometimes a dust particle would be in the absorption imaging beam path during the imaging pulse, producing strong circular fringes. There were also some circular fringes from the camera imaging system that seemed to vary slowly over time, possibly due to small temperature fluctuations. Usually, these subtracted out decently, but when they were changing rapidly enough, they caused some bad frames. There were several other sources of fringes that were mostly static, and may be from the lenses or the walls of the Hellma cell itself. Some of the fringes seemed to come and go as well. Since we were looking at a small signal in a bright light, relatively small changes in the light could have devastating effects on the final image. We corrected that by selectively culling pairs of images that significantly deviated from the average image. We would cull the same number of images from every frame in a data set so that every averaged image used in a given data set was averaged over the same number of images. The number we culled was selected so that there were no obviously bad frames in any of the averaged images, which we checked by visual inspection. This culling was largely for removing image pairs where the camera shutter did not activate properly, and so the number culled was largely dependent on how well the camera shutter was working.

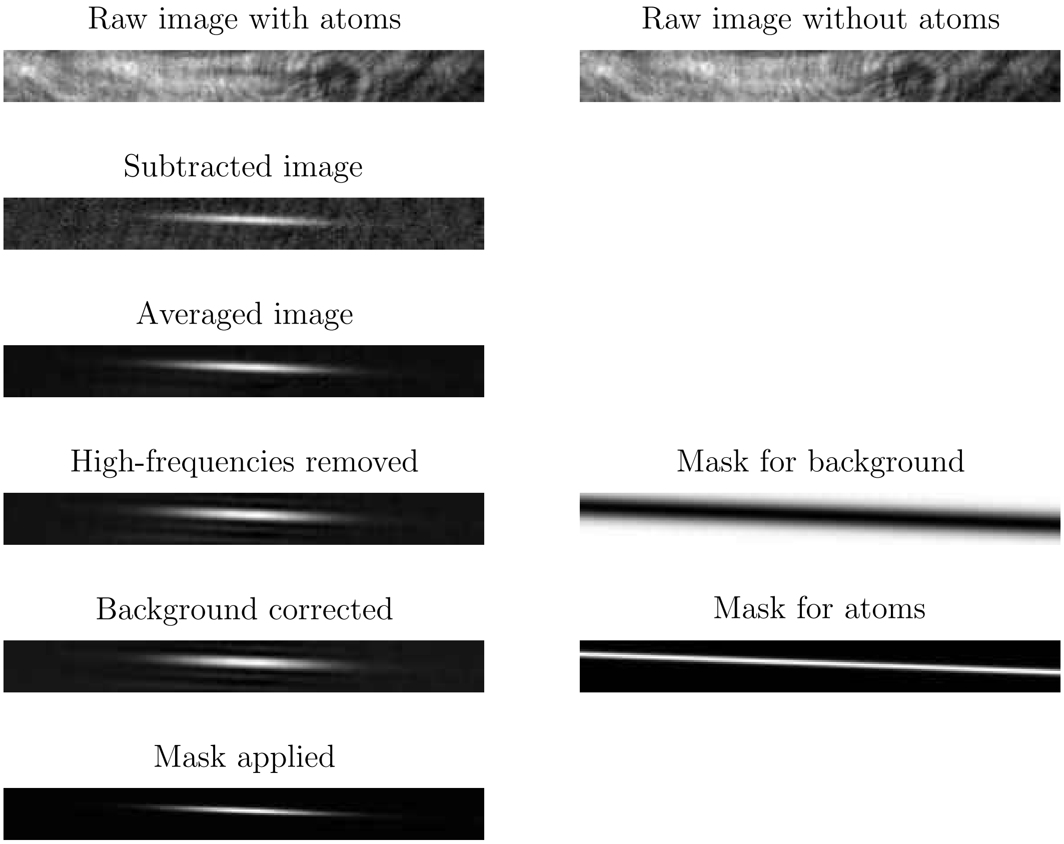


Figure 4.10. Processing one-way barrier images. We start with raw absorption images, which we subtract and normalize to remove the illumination and end up with something proportional to density of atoms. The result clearly shows atoms, but is noisy. We then average together many such subtracted images to reduce the noise. The excess high-frequency fringes are removed by taking the Fourier transform of the image, and forcing high-frequency contributions to zero, and then inverting the Fourier transform. The lower-frequency fringes are removed by doing a per-column background subtraction, using a wide mask to clip out the atoms. After applying a narrower mask to help remove the background further, the atoms are nicely shown. All the images here have had the contrast enhanced to cover the range from black to white; otherwise, the atoms would not be visible at a scale that showed the initial illumination pulse.

Once we removed images that were bad due to a misfired camera shutter, or particularly bad fringes, there were a few final processing steps to help remove fringes. First, to remove the near-vertical high-frequency fringes, we performed a Fourier transform to each row of the image, removed the high frequency components, and then transformed back. This process is shown in Figure 4.10. Using an image where there was no barrier, we would form a one-dimensional Gaussian mask around the atoms. For consistency, we used the same mask across every data set, until we moved either the camera or the dipole trap. The non-varying axis of this fit was oriented along the axis of the dipole trap (and was used to find the tilt of the dipole trap), so this mask only crops light away from the axis of the dipole trap. We would then use the inverse of a wider version of the mask to select out parts of the image that were definitely background. We then subtracted the average background value for each column from the averaged image. This tended to remove the larger fringes across the image, producing the final result shown in Figure 4.10.

Once we had a satisfactory image, we needed to perform just a few more steps to actually compute data from the image. We first apply the same mask used to differentiate background from the dipole trap region (but more narrowly restricted to the atoms). Usually, we used fluorescence imaging for determining the number of atoms in the trap, as the signal tended to be higher, but if we were using absorption imaging to count atoms, we would make sure the mask were large enough to not clip any atoms (or not use a mask at all). After applying the mask, we summed over the vertical dimension of the image (because it was close enough to perpendicular to the dipole trap axis), giving us a one-dimensional atomic density along the axis of the dipole trap. As we usually only wanted the number of atoms relative to the total, we did typically did not compare the fraction of light scattered by the atoms and compare with the scattering rate of non-saturated atoms to get absolute atomic numbers. These atomic density functions were the basis for a lot of the data we analyzed. The next most common step was to sum up the density function. If we summed the entire function, we could use the resulting number (proportional to the number of atoms in the trap) as a function of time to estimate trap lifetimes. However, since we know where the barrier beam is from aligning the barrier beams (you can see it in Figure 4.9), we can also sum up just the atoms on one side or the other of the barrier. This let us determine whether atoms were moving from one side of the barrier to the other. Pretty much all of the data we collected was in one of these forms.

We now present our main one-way barrier data in Figure 4.11, shown as one-dimensional atomics density plots as a function of time. The data is still a little noisy, so for this figure, we applied a 7-point second-degree Savitzky-Golay smoothing filter.[10]

Figure 4.11 has several columns, each showing a separate experiment. The "no barrier" column shows what happens when we load to one side for 5 ms and release, without turning on the barrier beams. Although the atoms appear to have been released on the left, we technically released them on the right and waited half a period for them to end up on the left, during which we made sure all atoms were in the transmitting state. The atoms oscillate about the focus of the dipole trap (the center of each column), slowly dephasing. As we will discuss in Section IV.5, this dephasing happens partly due to the anharmonicity of the potential, and partly as an effect of angular momentum. In contrast with the "no barrier" column, we present the "barrier" column, which shows the exact same experiment, except we turn the one-way barrier beams on around the time of the first image. At this point, atoms are on what we want to be the transmitting side. As we desired, the atoms transmitted through the barrier, but were unable to return, being trapped by the one-way barrier.

To differentiate our barrier from a single-pass barrier, we repeated the experiment, but started the atoms on the reflecting side (loading them on the left, waiting

[10] Savitzky-Golay is a rather simple way to smooth data that can preserve features (like widths of peaks) better than running averages. Essentially you fit a low-degree polynomial to a series of data points, and use the value of that polynomial at a given data point in place of the data point value. Fitting polynomials can be phrased as a linear algebra problem by projecting onto the basis of 1, x, x^2, etc. When phrased like this, if the x values (relative to the point you are replacing) are always the same, such as is the case for evenly spaced data, the projection becomes a matrix multiplication, where the matrix is the same for every data point (all that changes is the y values). Finding the value of polynomial then reduces to a running dot product with constant coefficients. In the limit of a 0-degree polynomial, this becomes a running average with equal weights, but for higher-order polynomials, this tends to better preserve peak widths.

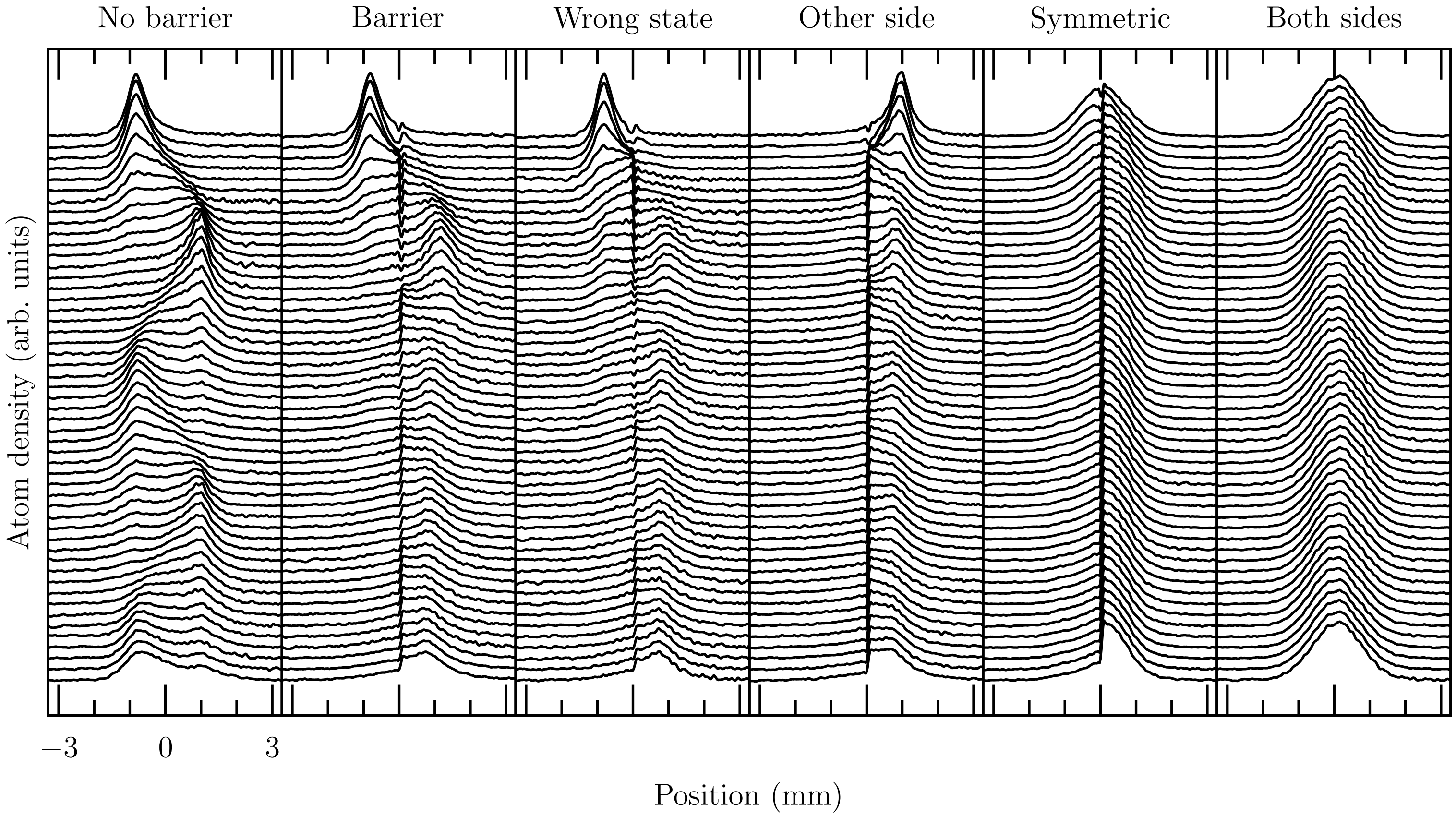


Figure 4.11. The distribution of atoms in the dipole trap as a function of time, showing the effects of the one-way barrier. The "no barrier" column shows what happens when we load atoms to one side of the dipole trap without the barrier. The "barrier" column shows the dynamics with the barrier beams on. In the "other side" and "wrong state" columns, the atoms start on the other side of the barrier in the former and in the reflecting state in the latter. In the "both sides" column, we let the atoms spread to fill the trap before turning the barrier beams on.

half a period until they were on the right), but still starting in the transmitting state. This is shown in the “other side” column in Figure 4.11. Here, the atoms never cross the barrier because they pass through the pumping beam of the barrier first, which puts them in the reflecting state. Then, when they hit the main barrier beam, they reflect instead of transmit. We can combine the effects of both the “barrier” and “other side” columns by performing a long load and letting the atoms equilibrate (the 110 ms load described above), and then turning the one-way barrier beams on. This has an extra benefit that we have more atoms from such a long load, and so the signal-to-noise ratio is higher. As shown in the “both sides” column of Figure 4.11, the atoms start with a distribution that is roughly symmetrical about the focus of the dipole trap. As shown in Figure 4.2, the atoms on the transmitting side can pass through the barrier. Atoms on the reflecting side, whether they start there or get there by passing through the barrier from the transmitting side, reflect off the barrier, and are trapped. The end result is nearly all the atoms end up on the reflecting side of the barrier, which makes a very visual demonstration of the one-way nature of this barrier. When we sum up the relative number of atoms on each side of the barrier, we get the time-series shown in Figure 4.12. We can see that the barrier is not perfect; there seem to be some atoms on the transmitting side of the barrier at the end of the experiment. Furthermore, there seems to be a slight decrease in the total number of atoms, much faster than we would expect from the lifetime of the dipole trap itself (which has a decay time on the order of 10 s, not 100 ms). Our simulations in Section IV.5 suggest that some of these atoms have not yet reached the barrier (as a result of having a large angular momentum), but that is not a full explanation. We believe these imperfections are the result of scattering events, which we will discuss more later.

One peculiarity of the one-way barrier is shown in the “wrong state” column of Figure 4.11. This column depicts what happens when we follow the same experimental procedure as the “barrier” column, but prepare the atoms into the reflecting state instead of the transmitting state. Since the atoms are in the reflecting state, we would expect the to bounce off of the one-way barrier. While some atoms do reflect on the first interaction with the barrier, as many atoms transmit. Of the atoms that bounce on the first interaction, almost all of them transmit on the second. The net result is all the atoms end up on the reflecting side again, although it takes a little longer than in the "barrier" column. As we pointed out in Figure 4.1, the main barrier beam is not equally detuned from the two ground states; it is much closer to resonance with the reflecting state. We will discuss this in more detail later, but essentially, this is because the maximum detuning from the ground states the barrier beam can have is not large enough to prevent an occasional interaction with a photon of light, altering the state of the atom. Later on, we will show some data reminiscent of when we were first trying the one-way barrier, while we were still locked to the rubidium 85 transition that was almost halfway in between the two rubidium transitions shown in Figure 4.3. The barrier simply does not work in that case, because there is just enough optical pumping to randomize the state of each atom as it crosses through the barrier. It only takes one absorption event to excite the atom, which will then decay into one or the other ground state, which is why even a very low scattering rate can have this effect.

While we will discuss ways to reduce this scattering effect later, we chose to keep our simple three-level atom version, with the same atomic levels, and pick the barrier detuning to reduce the effects of the scattering instead. By picking the asymmetric detuning shown in Figure 4.1, the main barrier beam actually scatters more light off of atoms in the reflecting state than the transmitting state, causing it to preferentially pump atoms into the transmitting state. This is probably why the so many atoms in the "wrong state" column of Figure 4.11 pass through on the first interaction with the main barrier beam. It also makes it quite likely that the atoms that do reflect are pumped into the transmitting state as they either leaving the main barrier beam after reflection, or as they enter it the second time, which explains why so many transmit on the second pass. Atoms on the reflecting side of the barrier do not transmit (often) because we intentionally placed the pumping barrier beam very close to the main barrier beam, so that as atoms reflect on that side, the pumping beam pumps atoms to the reflecting state faster than the main barrier beam pumps to the transmitting state. There is extra heating due to the extra scattering by having these two competing effects, but it allows the one-way

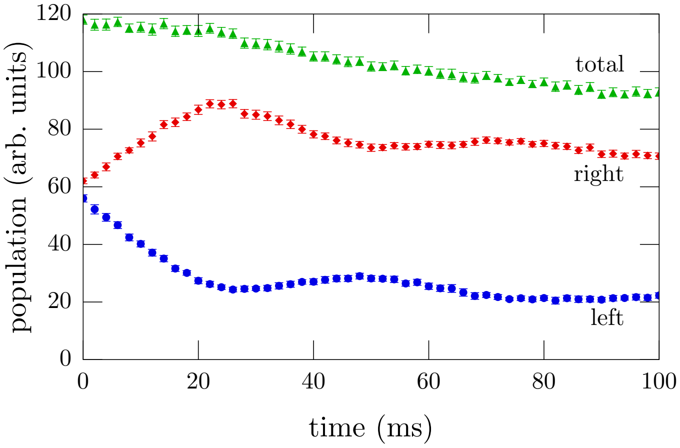


Figure 4.12. Populations on either side of the one-way barrier starting from a full load. We plot the estimated populations on either side of the barrier starting from a full 110 ms load, normalized to the approximate total population. The "both sides" column of Figure 4.11 shows the same data as position distributions of the atoms. Here, we see how the population on either side of the barrier start out approximately equal, but quickly skews to having all the atoms on the reflecting side as atoms pass from the transmitting side to the reflecting side, but the reverse process is blocked. We can also see a slight decay in the total number of atoms.

barrier to function. If this scattering is undesirable, there are other ways that might allow one to avoid these effects; however, as a rather unexpected benefit of this extra scattering, the one-way barrier now works even if you start the atom in the wrong (reflecting, as opposed to transmitting) state from the original idea.

Figure 4.13 illustrates how barrier-beam overlap helps the one-way barrier. Our usual separation of 34 (1) μm is marked on the plot, which shows the final, roughly steady-state population on either side of the barrier after 100 ms as a function of beam separation. The perfect one-way barrier would have all the population that had had the opportunity to cross the barrier on the reflecting side of the barrier. A very small fraction of the atoms may not make it to the barrier during the 100 ms, but that is rather negligible here. A barrier that leaks or reflects some atoms on the transmitting side will have some nonzero number of atoms on the transmitting side. Since a barrier might also simply eject atoms from the trap, things to look for are whetter the population on the transmitting side is small, and whether the total

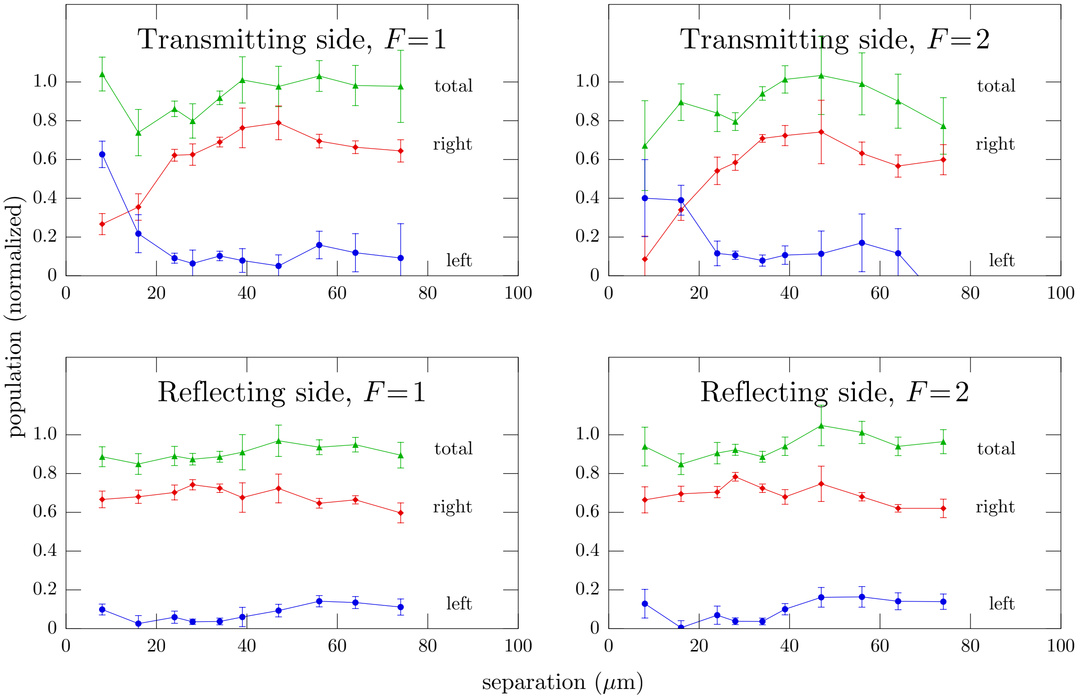


Figure 4.13. The effect of altering the separation of the barrier beams. Rather than showing an entire time-series such as in Figure 4.11 or Figure 4.12, we simply show the populations normalized to the initial number of atoms after 100 ms, when the populations have settled to roughly static values. This is equivalent to the last data points in Figure 4.12. Our typical barrier separation of 34 (1) μm is shown as the vertical gray column on the plots. The plots in the upper row show results if we start the atoms in the reflecting state, while the plots on the bottom show results for starting the atoms in the transmitting state. The plots on the left (right) show what happens when the atoms start on the left (right) side of the barrier when the barrier beams are turned on. We can see that making the beam separation much smaller drastically reduces the effectiveness of the barrier, while increasing the beam separation eventually reduces the effectiveness by a lesser amount.

population is unity (the number of atoms initially loaded). We see an optimum around 30 μm to 40 μm beam separation.

At our usual separation of 34 (1) μm, with beam waists on the order of 10 μm to 12 μm, the actual overlap is quite small, but apparently sufficient. As a rough approximation, we assume the pumping beam is resonant with a transition (which it would be if the dipole-trap beam did not Stark-shift the atoms out of resonance a little), with a peak intensity that is about 10 times the saturation intensity for rubidium. With a beam separation of about 3 $1/e^2$ intensity radii, at the peak of the main barrier beam, the pumping beam intensity is reduced by a factor of about e^{-6} to about 2% of the saturation intensity of rubidium. The main barrier beam is detuned by over 100 linewidths, and has a peak intensity of about 100 times that of the pumping beam, or about 1000 times the saturation intensity. For a far off-resonant beam, the scattering rate is proportional to intensity (divided by the saturation intensity) divided by 4 times the squared detuning (in linewidths), so the effective resonant intensity of this beam is about 3% of the saturation intensity. Those, to a rough approximation, the pumping beam can pump almost as fast as the main barrier beam even at the peak of the barrier beam, and is definitely dominant on the reflecting side, and not on the transmitting side.

So, according to this arguments, at significantly smaller separations than about 30 μm, the pumping beam will pump atoms to the reflecting state faster than the barrier beam pumps to the transmitting state over part of the *transmitting* side of the barrier. This will cause atoms to *reflect* off of the transmitting side of the barrier (and merely improve reflection off the reflecting side). According to this theory, at smaller separations we should expect to see the barrier become a strict barrier, keeping atoms on whichever side they start. We see this effect in Figure 4.13 where, at separations from a little over 20 μm and below, the populations on either side are highly skewed to whichever side the atoms started on. There also appears to be a large amount of loss when the atoms were initially on the transmitting side. This is probably because, at that separation, the pumping and main barrier beams compete at pumping the atoms to opposite states at the approximate turning point for many of the atoms. If the two beams compete in optical pumping, there will be more scattering, and if it happens near the turnaround point where the atoms spend a relatively large amount of time, this can result in significant heating, and hence possibly loss. There is also a separate loss mechanism, where, during a scattering event, atoms can temporarily combine into a molecule and gain a lot of kinetic energy in becoming bound, which ejects them from the trap. Since this mechanism, which we will discuss in Section IV.5, depends on the atoms being in the excited state, would also be greatly enhanced when there is more scattering of light.

We have mentioned that we want some overlap between the beams, because the main barrier beam slightly pumps atoms into the transmitting state. This is intended to help atoms transmit through the barrier, while the pumping beam helps them to reflect once they are on that side. If this overlap is necessary, we would expect to see the barrier start to leak as the beams get further apart and the overlap decreases. We do see a weak leakage in Figure 4.13 where, for large separation, regardless of which side of the barrier the atoms start out on, the final state has atoms on the supposedly transmitting side, although the effect is weak.

In addition to the beam separation, we also investigated the effects of the kinetic energy of the atoms hitting the barrier. By varying how far from the dipole trap center the atoms were loaded into the dipole trap, we could vary the kinetic energy the atoms would have when they fell onto the barrier. We stayed close enough to the harmonic region of the trap that the longitudinal oscillation period of atoms was always within a few milliseconds from 50 ms; as such, we did not need to alter the length of the half-period delay. As shown in Figure 4.14, changing the initial kinetic energy did not make much difference to the performance of the barrier. For an ideal, non-scattering barrier, the time-scale of the reflection is set by the ratio of the incoming kinetic energy to the barrier height, but the other dynamics are unaffected. Thus, altering the kinetic energy of atoms hitting the barrier is equivalent to varying the intensity of the main barrier beam. As we will describe later in this section when we estimate the number of scattering events, changing the intensity of the main barrier beam does have a weak effect on the number of scattering events, so this is close to, but not exactly, equivalent to changing the barrier height.

While we could have actually varied the main barrier beam power by changing neutral density filters at the barrier fiber input, we found it easier, and nearly equivalent, to change the initial loading position of the atoms. Changing the pumping beam power, however, was as simple as changing the control voltage of the AOM that controlled that beam power. We would expect that as we increased the power of the pumping barrier beam, the tail of the beam on the transmitting side of the

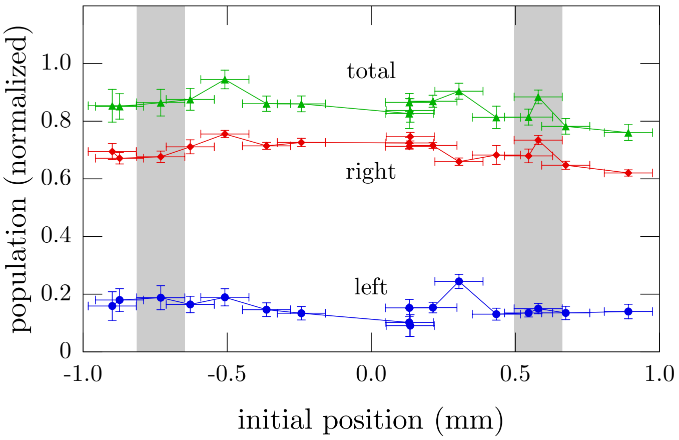


Figure 4.14. The effect on one-way barrier efficiency from altering the release point of atoms. As in Figure 4.13, we show the roughly steady-state value of the populations on either side of the barrier after 100 ms, except here this is shown as a function of initial loading position. The atoms are technically loaded into the dipole trap on the other side of the focus (and further back), and the location recorded here is where the center of mass of the atoms appear to be half an oscillation later, when we turn the barrier beams on. The usual effective starting locations are marked with vertical gray columns in the figure. As described before, this gives the other atoms from the MOT time to disperse, and allows us to pump the atoms into the proper state.

barrier would become strong enough to change the state of atoms to the reflecting state. In that event, the one-way barrier would become just a barrier that reflected all atoms. Were we to increase the pumping beam power even further, that beam would become a barrier in itself, without needing the main barrier beam. If we were to decrease the pumping beam power, we would eventually see it become too weak to reliably pump atoms into the reflecting state. In this case, we would see more atoms leaking through that barrier, since they were left in the transmitting state. Because the main barrier beam slowly pumps atoms to the transmitting state, with a very weak pumping beam, we would see no barrier effect at all.

We see effects like these in Figure 4.15. In this figure, we have plotted the final populations after 100 ms for a range of pumping beam powers that spans almost four orders of magnitude. Our usual pumping beam power is marked with a vertical bar. As we increase the beam power, we see the barrier continues to reflect

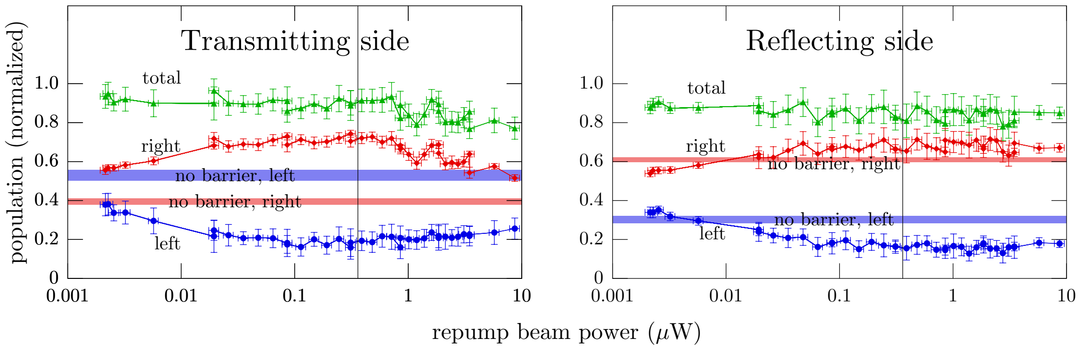


Figure 4.15. The effect on one-way barrier efficiency from altering the pumping beam power. As in Figure 4.13, we show the roughly steady-state value of the populations on either side of the barrier after 100 ms, except here this is shown as a function of pumping beam power. Our usual beam powers are marked with vertical gray columns. For comparison, the horizontal colored bars show the results of the same experiment without barrier beams. Without barrier beams, the asymmetry is simply a function of where the atoms were in there oscillation after 100 ms. For all the pumping beam powers shown here, there was enough of a barrier effect that any asymmetry shown is a result of the one-way barrier, and the populations were roughly steady-state. The decrease in asymmetry results in significant leakage of the barrier, but with enough damping to prevent significant continuing oscillation of the center of mass.

atoms that start on the reflecting side just as well as ever, but when the atoms are released on the transmitting side, we see the one-way barrier starts keeping more and more atoms on the supposedly transmitting side. We also see increased loss as the pumping beam power is increased, suggesting that competitive scattering between the main barrier and pumping beams is heating atoms. As we decrease the pumping beam power, we see the one-way barrier efficiency decreasing without atom loss, regardless of which side of the barrier the atoms start out on. This is probably because the atoms are less and less kept in the reflecting state, and so can pass through the barrier.

The final barrier parameter we varied is the detuning of the main barrier beam, with some results shown in Figure 4.16. The detunings we used are shown in Figure 4.3 and listed in Table 4.1. The typical detuning is the one we used for almost all the data we took, as the barrier functioned well with that frequency and it was easy to lock our lasers to that frequency, since there happened to be a rubidium 85 transition there. The middle detuning is the one we first tried, as it is nearly symmetrically detuned from transitions in either the reflecting or transmitting state, and also is convenient to lock to, as there is a rubidium 85 transition there as well. The rubidium 85 transitions are too close together to easily resolve, resulting in one effective transition to which we locked our lasers. We believe this is close to the rubidium 85 $F = 2 \longleftrightarrow F' = 2$ transition, but we will report it as the $F = 2 \longleftrightarrow F'$ transition. Counter-intuitively, the middle detuning worked poorly, as demonstrated in Figure 4.16. Since we checked to the red of the typical detuning, we also checked to the blue. Nothing in our rubidium vapor cells has any transitions between the typical detuning and the reflecting-state transitions, save more of the same rubidium 85 $F = 3 \longleftrightarrow F'$ peaks used for the typical detuning, but these were too close. As such, as opposed to locking to main barrier beam to a particular frequency, we let it drift. To make sure the laser neither drifted too far from the frequency nor mode-hopped to a different frequency, we monitored its spectrum with a Fabry-Pérot cavity. We would first manually detune it just a little to the blue of the reflecting-state transitions, as seen by the saturated absorption spectroscopy setup that we used to lock to the rubidium 85 lines. Then we would watch the spectrum of the Fabry-Pérot cavity which had pickoffs of both the MOT trapping laser and the main barrier beams coupled into it. By tuning the barrier beam to appear halfway between two MOT beam peaks in the Fabry-Pérot spectrum, we made sure it was about 0.75 (5) GHz (half the free spectral range of the cavity) detuned from the main MOT trapping line (which happens to be the closest reflecting-state transition). While taking data, we would frequently monitor this spectrum, recording it with an oscilloscope, so we would know when the laser drifted too far from our intended detuning and could retake that data. By fitting two sets of evenly spaced Lorentzians to the spectrums (one for the MOT laser and one for the barrier laser), and using the known free spectral range as a calibration, we could determine the relative detuning of the two lasers relatively accurately, modulo the free spectral range. While taking data, we allowed the detuning to drift by about 0.05 GHz, which is where our assumed error

	Detuning from reflecting state:	Detuning from transmitting state:	Lock method:
typical	$1.05\,(5)$ GHz	$-5.29\,(5)$ GHz	^{85}Rb $F = 3 \longleftrightarrow F' = 3, 4$
middle	$3.97\,(7)$ GHz	$-2.37\,(7)$ GHz	^{85}Rb $F = 2 \longleftrightarrow F'$
near	$0.75\,(5)$ GHz	$-5.59\,(5)$ GHz	Fabry-Pérot

Table 4.1. The various detunings we used when varying the detuning of the main barrier beam. All values are blue of the reflecting-state transitions and red of the transmitting-state transitions, and multiple excited states are listed because we technically detuned to the crossover peak midway between the two given transitions. Detunings are given relative to the closest (highest frequency) of the reflecting-state transitions ($F = 2 \longleftrightarrow F' = 3$) and the closest (lowest frequency) of the transmitting-state transitions ($F = 1 \longleftrightarrow F' = 0$). These detunings are shown graphically in Figure 4.3.

comes from. The chances of a mode hop which changed the frequency by an even multiple of the free spectral range are small. Since such a hop should would either make the barrier perfectly reflecting (blue-detuned for both ground states), near the transmitting state transitions, very far detuned from all states, or put the detuning somewhere near the middle detuning, we assume we would notice the effects of such a hop, as all make noticeable difference in the evolution (the middle detuning difference can be seen in Figure 4.16).

Using the three detunings described above, and in all cases varying the intensity of the main barrier beam so that the height of the reflecting barrier should be similar in all three cases, we turned the barrier beams on when the atoms were on the reflecting side of the barrier, and observed the populations on either side of the barrier for 500 ms (5 times longer than many of our data sets). The results are plotted in Figure 4.16. All the detunings appear to keep atoms from penetrating the barrier to the transmitting side. We see that, while none of the detunings are perfect, the loss rate at the middle detuning is much higher than either the near or typical detunings.

The results in Figure 4.16 suggest something about the deficiencies of this one-

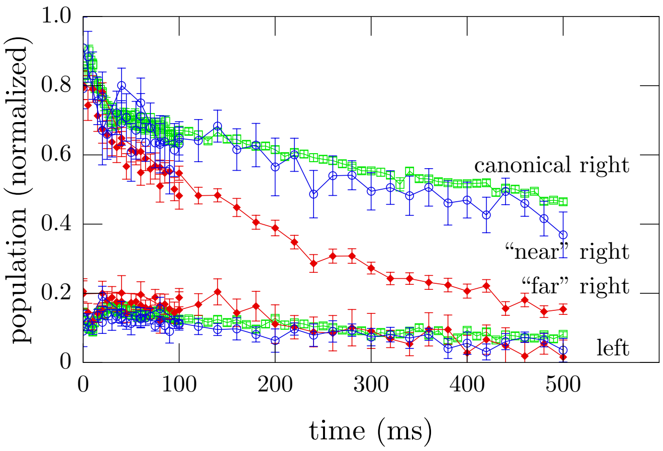


Figure 4.16. The effect on one-way barrier efficiency from altering the main-barrier-beam detuning. Here we show the populations on either side of the barrier as a function of time, over a longer period of time (500 ms) than we show in several other figures (100 ms). The detunings are shown in Figure 4.3, with values given in Table 4.1. This plot shows the lifetimes of the of atoms on the reflecting side of the barrier, for each of these detunings. For the near and typical detunings, which tend to pump atoms into the transmitting state, the curves are similar. The middle detuning demonstrates a much shorter lifetime.

way barrier. Since the barrier height should be similar in all three cases, there is no reason to suspect that the conservative effects of the light beam are at play here. That leaves the non-conservative effects of the light on the atoms; that is, scattering. Each scattering event gives a small momentum kick to the atom, effectively heating up the atoms. A scattering event also effectively randomizes the state of the atom. The heating effect is a bit of an issue, but if it were the main issue, the near detuning would have the largest scattering rate. The fact that the majority of the atoms stay on the reflecting side of the barrier suggests that the atoms spend most of their time in the reflecting state, perhaps helped by the pumping barrier beam, so the near-detuned main barrier beam should scatter more light off atoms than any other detuning. Since the near detuning does not have substantially more loss than the other detuning, we conclude that the heating effect directly from scattering light is a small effect here. The remaining effect is the resulting state from the atoms. The typical detuning and the near detuning should tend to pump atoms into the transmitting state, but not the other direction. However, the middle detuning, being roughly equally detuned from the reflecting states as the transmitting states, is about as likely to pump from the reflecting states to the transmitting states.

We believe this difference in relative pumping rates is the reason the middle detuning works poorly here. The middle detuning is on the order of twice as far detuned from the transmitting-state resonances as the typical detuning. To have the same barrier height, since barrier height is proportional to intensity and inversely proportional to detuning (Equation (II.9)). Since scattering rate is proportional to intensity, and inversely proportional to detuning squared (Equation (II.10)), the scattering rate is only about half as much for the middle detuning as for the typical detuning. Thus, if the typical detuning tends to flip the state of atoms as they are leaving the barrier approximately 10% of the time, and does nothing to them afterwards (since they are the too far out of resonance), there is something like a 5% chance that the middle detuning will flip the state of the atom. Squaring that gives a not-insignificant chance that the atom will be This would not matter if the middle detuning were far enough detuned to prevent any scattering at all, but, in the presence of sufficient scattering, a preference to pump only one direction (reflecting to transmitting) works better.

We also tried a variant of the polarizations of the beam. Normally, the main barrier beam was linearly polarized parallel to the surface of the optical table and the axis of the dipole-trap beam, while the pumping barrier beam was linearly polarized and perpendicular to the dipole-trap beam (vertically polarized). We wanted to change the polarization in the easiest way possible, and we had several options. We could have rotated the outputs of the barrier-beam fibers, but we would then need to find the MOT with the beams again, as the beams do not exit the fiber couples exactly on axis; we could have rotated the inputs of the barrier-beam fibers, but we would then need to recouple the fibers; or we could have inserted a waveplate somewhere in the beam path, which would probably require some realignment. We elected to rotate the main barrier beam fiber input by 90°, and re-couple that fiber. We did not want to change the pumping beam polarization, as we had spent quite a bit of time figuring out the mapping between AOM control voltages and pumping beam power within the cell, which was difficult because the relationship is nonlinear, and the powers we used were low enough that we needed to measure them very carefully. While we could have inferred the correct AOM control voltage once we recalibrated the transmission through the prism pair, we would still have wanted to verify it, which would require a few sensitive power measurements. Compared to re-calibrating the pumping beam power, or possibly having to find the MOT with the barrier beams again, coupling a fiber input seemed quite simple, so we elected for that option, as there were no waveplates readily available. Re-calibrating the power would not have been too complicated, but definitely more complicated than recoupling a fiber input. We ended up rotating the main barrier beam fiber input coupler, so that both barrier beams were linearly and vertically polarized.[11] Once we recalibrated the main barrier beam power, which just took a few transmission measurements through the prism pair, we took enough data with these polarizations to verify that the behavior of the barrier did not change in a very noticeable way. We conclude that the barrier is probably not very sensitive to beam polarization.

In summary, we have demonstrated that our optical one-way barrier works under a rather wide range of conditions. Scattering is an important effect, and our particular choice of transitions prohibits us from reducing that scattering to a negligible level. We do not consider this to be a fundamental limitation. Scattering could be greatly reduced by choosing atomic states and transitions that do not share a common excited state. This could be done, for example, by using polarization-sensitive transitions and magnetic sublevels. Another alternative is to use more than just three levels of the atom, particularly utilizing long-lived metastable states and exciting transitions available only to those states for the barrier [20].

[11] As mentioned previously, these polarizations were reported incorrectly in one of our publications [20]. In that paper, horizontal (parallel to the dipole-trap axis) and vertical (perpendicular to the dipole-trap axis) polarizations were all switched. The pumping barrier beam was always vertically polarized (not horizontal, as reported in the Physical Review A article), the main barrier beam was usually horizontally polarized (not vertically), and we rotated to to match the horizontal pumping barrier beam (not vertical).

Simulating our Optical One-Way Barrier

Using a rather simple semi-classical simulation, we were able to successfully model the behavior of our one-way barrier. The atoms were assumed to be non-interacting, moving within the main trapping laser, and affected by the two barrier beams. The atoms themselves are modeled as being in one of two states, trapped (corresponding to the $F = 2$ ground state of rubidium) and untrapped (the $F = 1$ ground state). Quantum-mechanical coherences are not explicitly used by the simulation, but are used to compute the magnitude of effects the lasers have on the atoms. All the laser beams are treated as affecting the atoms in two ways. First, they produce conservative potentials in which the atoms move. Second, they scatter light of the atom. Each scattering event is random, and imparts a two-part momentum kick to the atom, as well as a random chance of changing the state of the atom. Both the scattering probability and the potential seen by the atoms are dependent on the state the atom is in. We will discuss each of these parts in greater detail.

The simulations break real time into very small time steps. There are three time-scales in the simulation. The main trapping beam gives a very weak confining potential along the beam axis, which results in a rather slow motion of the atoms along the length of the trap. Motion along this direction takes place on the time-scale of milliseconds, with a single oscillation happening on the order of fifty milliseconds. The radial direction, along with the narrow waists of the barrier beams, present very steep potentials, which can impart strong forces on the atom. Since the relevant length scales of these potentials are tens of microns, as opposed to the millimeter-scale Rayleigh length of the main trapping beam, the time-scales for crossing these beams is on the order of tens of microseconds. Also, while within the barrier beams, the atoms may experience scattering effects which have maximum rates on the order of tens of megahertz. These events therefore take place on the tens of nanosecond time-scales. Outside of the barrier beams, the only effect on the atoms is from the main trapping beam, which is radially symmetric. By conserving radial angular momentum, we can reduce the problem from three dimensions to two, with millisecond time-scales. Only within the barrier beams do we need to worry about shorter time-scales, with the fastest time-scales only an issue when the atoms are in states that are likely to scatter. This setup would have been a great one to develop an adaptive time step, but because computational resources were more readily available to us (even a standard desktop was sufficient) than our time, we chose to use fixed time steps on the order of nanoseconds (5ns), which is fairly fast compared to even the fastest time-scales of the system.

To integrate the motion of the atoms through the conservative potentials, we used fourth-order Runge-Kutta. The Runge-Kutta methods are quite standard, and one can find them in many references on numerical analysis and methods or differential equations [89–91]. The basic idea is to find a function knowing only its derivative. This is common in classical physics, where we have position and momentum. The derivative of the position is the momentum (scaled by the mass), and the derivative of the momentum is given by the potential. Basically, the idea is to assume the answer is a rather smooth curve, and approximate the derivatives of that curve to develop a Taylor expansion of the polynomial. That Taylor expansion is used to get the next point on the curve, and the process is repeated.

The cleverness behind the Runge-Kutta methods is that instead of actually having to evaluate higher-order derivatives of the function directly, they are approximated by evaluating the first derivative at multiple points. The simplest version, known as the Euler method, involves simply adding the first derivative (multiplied by a time step) to the state with every iteration. This fails in the case of a circular orbit, because the first derivative *always* points towards the outside of the circle, and so no matter how small the time step is, the computed orbit always diverges. A simple way to fix this is to try to evaluate the derivative after half a time step, which gets closer to the right answer. That is the basis behind higher-order Runge-Kutta methods.

The standard (but certainly not unique) fourth-order Runge-Kutta method we use is as follows. We have a state vector $\overrightarrow{x}(t)$, and wish to approximate $\overrightarrow{x}(t + \Delta t)$. We take the first derivative ($\overrightarrow{k_1}$) and use that to approximate the state after *half* a time step, where we evaluate the first derivative again ($\overrightarrow{k_2}$). Using $\overrightarrow{k_2}$, we re-evaluate the half-time-step state, and compute another derivative ($\overrightarrow{k_3}$). Finally, using $\overrightarrow{k_3}$, we compute an approximate of the full time step state, and compute our last derivative there ($\overrightarrow{k_4}$). We then use a weighted sum of the four derivatives we computed to approximate the state after a full time step. The weights are chosen to cancel as many terms of the Taylor expansion as possible.[12] Here is the formula, suppressing the time-dependence of the state ($\overrightarrow{x}$ means $\overrightarrow{x}(t)$):

$$\begin{aligned}
\overrightarrow{k_1} &= \overrightarrow{F}(\overrightarrow{x}, t) && \text{Current derivative} \\
\overrightarrow{k_2} &= \overrightarrow{F}\left(\overrightarrow{x} + \overrightarrow{k_1}\frac{\Delta t}{2}, t + \frac{\Delta t}{2}\right) && \text{Derivative at first estimate} \\
\overrightarrow{k_3} &= \overrightarrow{F}\left(\overrightarrow{x} + \overrightarrow{k_2}\frac{\Delta t}{2}, t + \frac{\Delta t}{2}\right) && \text{Derivative at second estimate} \\
\overrightarrow{k_4} &= \overrightarrow{F}\left(\overrightarrow{x} + \overrightarrow{k_3}\Delta t, t\right) && \text{Derivative at third estimate} \\
\overrightarrow{x}(t + \Delta t) &\approx \overrightarrow{x} + \frac{\Delta t}{6}\left(\overrightarrow{k_1} + 2\overrightarrow{k_2} + 2\overrightarrow{k_3} + \overrightarrow{k_4}\right). && \text{(IV.2)}
\end{aligned}$$

In our simulation, we assume all the beams may be approximated as far-off-resonance beams interacting with two-level atoms. This results in conservative potentials proportional to the local beam intensity, and inversely proportional to the detuning of the beam from the atom, as given by Equation (II.9). The potentials from different beams simply add. In reality, the beams are not all that far from the atomic resonances, and so we also add a scattering effect between time steps of the simulation. The scattering effect is computed separately for each beam, with a scattering probability (the rate multiplied by the time step) proportional to the local beam intensity and inversely proportional to the beam detuning, as given by Equation (II.10).

Since the atoms are not two-level atoms, we invoke a separation of time-scales argument to avoid performing quantum mechanical computations with each time step. The atomic energy levels are separated by at least megahertz, if not hundreds of megahertz or gigahertz (multiplied by Planck's constant), so atomic coherences evolve on the microsecond to nanosecond ranges. The atomic excited states have

[12] Demonstrating this is rather straightforward, but requires simplifying many terms, and so we omit a discussion of that.

lifetimes on the order of tens of nanoseconds. These time-scales are much faster than any other time-scale in the problem, so we make the approximation that the quantum state of each atom is continually in a steady state for the given beam intensity. Furthermore, since the beams are not phase locked with each other, and do not have large overlaps, we ignore the possibility of any interference between beams or their effects on the atoms.

This approximation allows us to simply keep track of which ground state the atom is in (representing a collapse to a known state after a scattering event). Given that, we can account for the multiple levels in the atom by computing an effective detuning taking into account interference effects and the fact that different levels have different couplings to the laser beam. This gives us a state-dependent effective detuning for both the conservative potential and the scattering rate. The same computation also gives us the probability of a change to the other ground state when the atom does scatter a photon. The results of the computation are summarized in Table 4.2. Because the potential and scattering rates have different detuning dependencies, the effective detunings for the two effects are different.[13] Furthermore, due to the different couplings of the states, the differences in detunings for the two ground states do not add up to the frequency splitting of the two ground states. Neither of these discrepancies are particularly significant. We note that the pumping beam is intended to be on resonance with a transition for the $F = 1$ ground state. However, the trapping beam shifts the atoms out of resonance, an effect mentioned in Section II.3. The exact value is position dependent, so the values quoted for detunings in Table 4.2 are a decent approximation of the total detuning that we used in our simulations. The pumping beam produces a very high scattering rate for much of its width, even with the detuning, and so the exact value is not too important as the atoms are quickly pumped to the reflecting state with or without a detuning. The detuning is small enough that there is no potential effect for atoms in the $F = 1$ ground state (scattering dominates). Since the pumping beam is weak and far-detuned for atoms in the $F = 2$ ground state, we found that the potential

[13] The detunings are close enough that we used a single value for both scattering-rate and potential computations in our simulations.

Barrier beam

	Transmitting state $F = 1$	Reflecting state $F = 2$
Potential detuning	$\Delta f = -5.413$ GHz	$\Delta f = 1.123$ GHz
Scattering detuning	$\Delta f = -5.412$ GHz $\mathcal{P} = \frac{5}{18}$	$\Delta f = 1.117$ GHz $\mathcal{P} = \frac{1}{6}$

Pumping beam

	Transmitting state $F = 1$	Reflecting state $F = 2$
Potential detuning	$\Delta f = -16$ MHz	$\Delta f = 6.637$ GHz
Scattering detuning	$\Delta f = -16$ MHz $\mathcal{P} = \frac{1}{2}$	$\Delta f = 6.636$ GHz $\mathcal{P} = \frac{1}{2}$

Table 4.2. The effective detunings used in the barrier simulations. Detunings both for computing dipole potentials and scattering rates, for atoms in either states, are shown. Also included are the probabilities for changing state when the atom scatters light.

was negligible for those atoms as well, and so we simply disabled the computation of the pumping-beam potential in our simulation. The shift in resonant frequency due to the trapping beam was relatively small compared to the other detunings, and so we ignored the effect for those.

The intensity profiles of the beam are computed assuming a standard Gaussian beam shape, commonly derived in standard optics textbooks [92]. If the beam is assumed to propagate along the z-axis with a focus at $z = 0$ and an integrated power of P, then the intensity I is given by:

$$I(x, y, z) = \frac{2P}{\pi w_{x0} w_{y0}} \frac{e^{-2\left(\frac{x}{w_x(z)}\right)^2} e^{-2\left(\frac{y}{w_y(z)}\right)^2}}{1 - \left(\frac{z}{z_0}\right)^2} \tag{IV.3}$$

$$w_{x,y}(z) = w_{x0,y0}\sqrt{1 - \left(\frac{z}{z_0}\right)^2} \tag{IV.4}$$

$$z_0 := \frac{\pi w_{x0} w_{y0}}{\lambda}. \tag{IV.5}$$

This is normally written with the two waists ($1/e^2$ intensity radii) at the focus (w_{x0} and w_{y0}) equal, but our barrier beams are asymmetric, and so we need the two waists to be separate. The Rayleigh length, z_0, is the length scale of the focal region, and λ is the wavelength of the light.

Our simulation consists of three beams. The main trapping beam defines the z-axis, with the focus at the origin, with the propagation being in the $+z$ direction. Gravity (which is included in the simulations) defines the $+y$ direction, and the barrier beams are assumed to propagate in the $+x$ direction.[14] In our simulations, the main trapping beam typically had a power of 9.34 W (in some, we simply used 10 W), which we calculated to be the approximate power of the trapping beam that actually transmitted into the vacuum chamber. The maximum scattering rate of this beam is on the order of 3 s^{-1}, and so we ignore scattering off of the trapping beam. The simulated trapping beam is symmetric, with a beam waist of 30.9 μm, and operating wavelength of 1090 nm. With these numbers, the Rayleigh length works out to be 2.8 mm and the trap depth works out to be $k_\mathrm{B} \times 0.9$ mK. The barrier and pumping beam are assumed to cross the trapping beam perpendicular to the axis, a distance of 17 μm to each side of the focus (for a separation of 34 μm). The main barrier beam has a beam waist of 11.5 μm along the trapping beam axis, and 80 μm perpendicular to that, with a total power of 40 μW. The pumping barrier beam has a beam waist of 13 μm along the trapping beam axis, and 60 μm perpendicular to that, with a total power of 0.36 μW. The detunings of

[14] In more complicated simulations, we also allowed for the fact that the barrier beams do not propagate perfectly perpendicular to the dipole trap beam, but the effect of this was small, and the results we discuss in this section do not include that effect.

these beams are given in Table 4.2. With these parameters, the barrier beam has a peak potential of $k_{\mathrm{B}} \times 0.22$ mK ($k_{\mathrm{B}} \times -0.05$ mK $\times k_{\mathrm{B}}$) and a maximum scattering rate of 1.5×10^5 s^{-1} (6.6×10^3 s^{-1}) for atoms in the reflecting (transmitting) state. As mentioned, we disable the pumping beam potential. The maximum scattering rates for the pumping beam are 47 s^{-1} (5.6×10^6 s^{-1}) for atoms in the reflecting state (transmitting) state.

For each time step of the simulation, we compute the force on each atom using an analytic formula for the gradient of the potentials from the three laser beams (according to the state of the atom), and iterate forwards in time using fourth-order Runge-Kutta. We then pick a random number for each atom and each beam and use it to determine if that atoms scatters a photon from that beam, using the state-dependent local scattering rate multiplied by the time step as the probability of a scattering event. If the atom scatters a photon, we add one photon-recoil of momentum to the atom in the beam direction to simulate the kick from absorption of the beam (we pretend the beam is has no spread in propagation direction for this). We then add in a randomly directed photon-recoil of momentum, with a dipole-radiation pattern, to simulate spontaneous emission of the atom excited by linearly polarized light. Using the probability of a state change given in Table 4.2, we then randomly change the state of the atom to simulate a possible decay into the other ground state. After this, we proceed to the next time step.

We can now try to match our simulation with our experiments. The atoms are initialized to a close approximation of the initial conditions of the atoms in the experiments. They start with a Gaussian distribution 0.9 mm from the focus of the trapping beam, with a position-independent thermal momentum distribution corresponding to 100 μK. The longitudinal width of the positional distribution corresponds to the approximate measured size of the atom cloud in our MOT, while the radial directions are compressed slightly, to reduce the number of atoms in the simulation that simply fall out of the trap. In the actual experiment, atoms load from the MOT into the dipole trap continuously through the loading time, although probably not at a constant rate. Since the dipole trap shifts atoms out of resonance with the MOT trapping beams, the atoms that enter the trap earlier start their oscillations earlier. We approximate this effect by freezing the atoms initially and releasing them at random times during the MOT loading period in the simulation. Each atom is started in the same ground state to which we attempted to initialize the atoms in the experiment. We then let the atoms evolve in the simulation for 27 ms before enabling the barrier and pumping beams in the simulation, which includes a 5 ms loading time and another 22 ms for the atoms to travel to the other side of the trap.

While we have also compared position distributions from both the simulation and experiment, we found it simpler to compare the relative numbers of atoms on each side of the barrier, as a function of time. The position distributions are a decent match, as might be inferred from the comparison of numbers to each side of the barrier, shown in Figure 4.17. We feel the simulations closely match the results of the experiment.

Comparisons of simulations and experiment demonstrated two effects which we had not expected. We knew that the period of atoms in the trap (about 50 ms) were longer than the period we would expect from the harmonic center of the trap (about 42 ms), and dephased within a few oscillations. We thought these were largely effects of the anharmonicity of the trap, as the atoms are on the order of a Rayleigh length away from the trap focus, where the potential is noticeably non-harmonic. However, comparisons of our experiment with our simulation that, while this effect did contribute, there was another contribution. The results shows in Figure 4.18 shows a similar comparison as in Figure 4.17, but without the barrier beams enabled. There are two simulations. In the simulation given by the solid curve, the initial conditions are set up to match our experimental conditions. In the dashed curve, the initial conditions are the same, except we set all the radial positions and momenta to zero.

As seen in Figure 4.18, this change in initial conditions had a drastic effect on the result. Since both setups had the same longitudinal spread, to lowest order they should match on the anharmonicity of the trap. We see that anharmonicity does have an effect, as the trap period is still larger than what we would expect from the harmonic part of the trap. However, the initial condition with radial

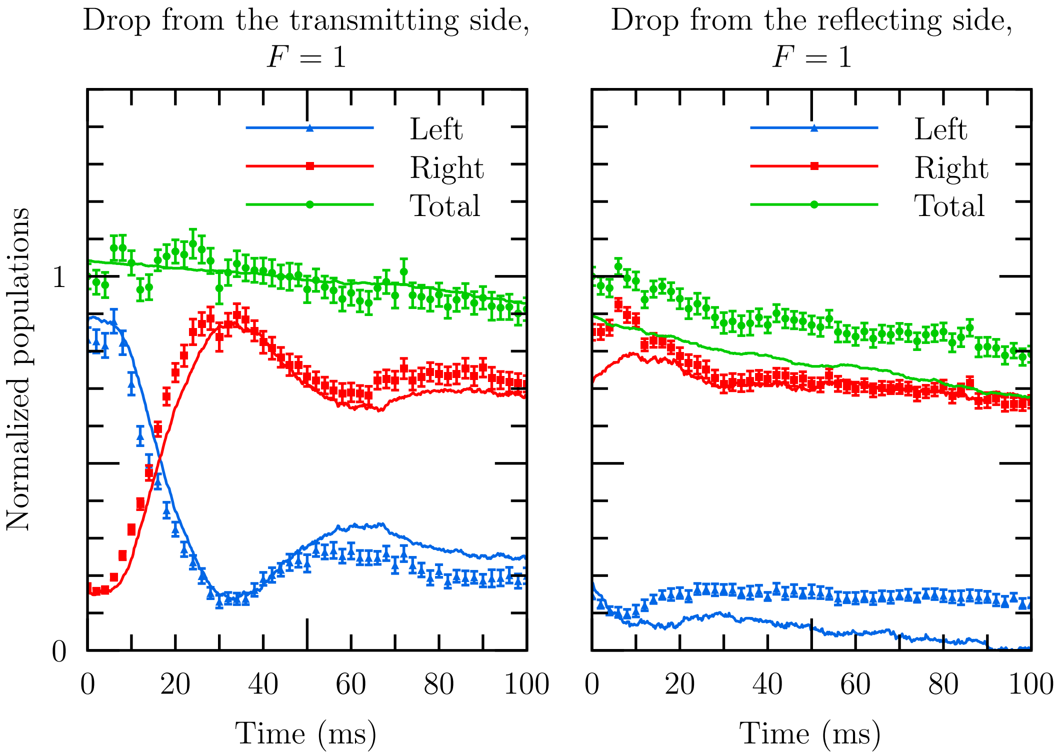


Figure 4.17. A comparison of simulation to typical experiments. Rather than attempt a full comparison of atomic distribution (which matched relatively well), we typically compared only the relative numbers of atoms on each side of the barrier. The individual points represent data from the experiment, while the line is the simulated data. The green points and curve show the total number of atoms still in the trapping region, while the red and blue data represent the number of atoms on the left and right sides of the barrier, respectively. Data for atoms starting on both sides of the barrier are used.

motion has a larger period still, which suggests a coupling between radial motion and longitudinal motion. Since the potential does not decouple into the product of radial and longitudinal factors, some coupling is inevitable. However, the potential is radially symmetric allowing us to invoke conservation of angular momentum to reduce the equations of motion to a radial and a longitudinal component, as opposed to the full three dimensions. With these equations, which we discuss later, we see that angular momentum effectively flattens out the trap, and can even make the central focus of the laser effectively repulsive (although this requires a rather large

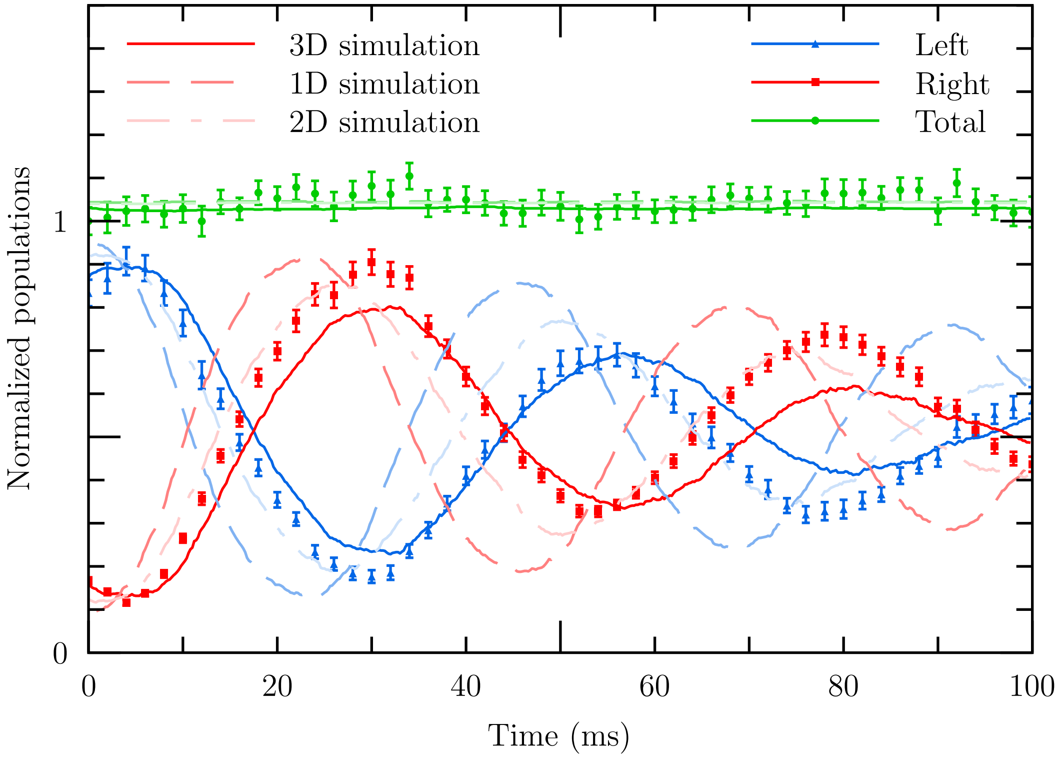


Figure 4.18. A comparison of simulations with different initial conditions. Here we compare various simulations with data taken from an experiment without the barrier beam turned on (the "no barrier" column in Figure 4.11). The points represent data from the experiment (relative numbers of atoms to each side of where the barrier would be). The solid curve shows the simulation results using our typical initial conditions. The fainter, dashed curve shows the results if all the atoms are confined to the axis of the main trap, while the faintest, dot-dashed curve shows the results if the atoms are confined to the plane defined by the axis of the trap and gravity. The green points and curves show the total number of atoms, while the red and blue data show the numbers to the right and left of where the barrier would be (the focus of the main trapping beam), respectively.

initial kinetic energy compared to the average kinetic energy of atoms in our trap). This is because the confinement of the focus requires atoms with angular momentum to make a tighter, faster spiral near the focus of the trap. If the angular momentum is large enough, the kinetic energy required to pass through the focus can be larger than the depth of the trap, effectively preventing trapped atoms from entering the focus.

Anharmonicity results in dephasing because, unlike in a harmonic trap, atoms with different starting conditions have different periods of oscillation.[15] This angular momentum effect amplifies this, as not only do atoms have different velocities and initial positions, they also have differing values of angular momentum, which causes them to see different effective longitudinal potentials. Differing potentials results in different periods for the atoms as well, increasing the rate of dephasing as the atoms oscillate back and forth.

The second effect we noticed showed up when we attempted to use the simulations to check the lifetimes of atoms within the trap. Over the typical 100 ms, the effect is not very noticeable, but, as shown in Figure 4.19, over the second time-scale the effect is dominant. The simulations, as we have described them so far, cannot predict even the magnitude of the lifetime of atoms in the trap. As the lifetime of atoms in the trapping beam without the barrier beam and pumping beam can be much longer, whatever was missing in our simulations obviously depended on the barrier beams. However, we had accounted for every single-beam, single-atom effect we knew of, and something else was clearly ejecting atoms from the trap. Multi-beam effects seemed unlikely to us, as those usually require the sum frequencies of the beams to nearly match some transition of the atoms, but the beam frequencies we had do not match any of the transitions in rubidium. We did know of various multi-atom effects, which are discussed in many papers [93, 94]. In particular, we referenced a paper from Carl Wieman's group, which discussed these effects in rubidium 85 atoms in a dipole trap, with a setup very similar to ours [61].

[15] Technically, due to the three dimensions coupling in this potential, the orbit of an atom is not periodic. However, the atoms do tend to reach a similar point along the axis of the trapping beam, and the time it takes for that is what we mean by the "period" of an orbit.

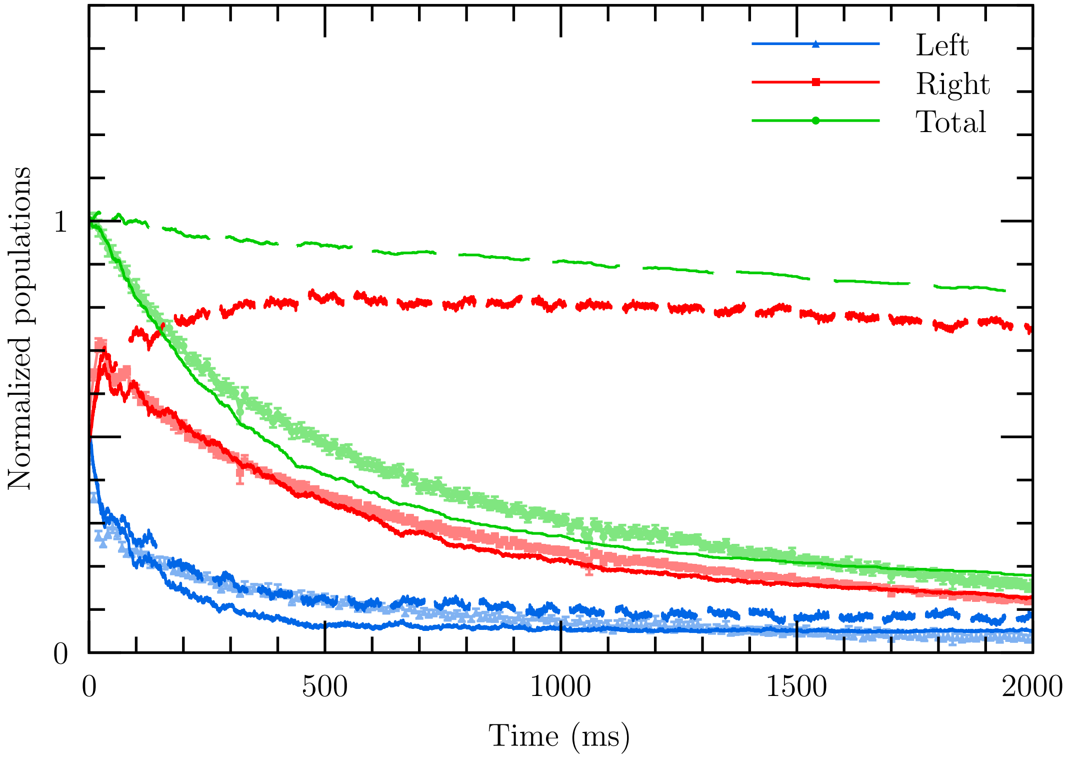


Figure 4.19. A comparison of simulation lifetimes to experiment. For these plots, the atoms start on the reflecting side of the barrier, and we carry the experiment out over a longer-than-usual period of time to measure the lifetime of the atoms with the barrier. The points represent data from the experiment (relative numbers of atoms to each side of where the barrier would be). The solid curve shows the simulation results including the light-assisted collision mechanism. The dashed curve shows the results without that mechanism. The green points and curves show the total number of atoms, while the red and blue data show the numbers to the right and left of where the barrier would be (the focus of the main trapping beam), respectively.

There are several multi-atom effects we could consider. Pairs of rubidium atoms could temporarily combine into a molecule (aided by light from the trapping beam) with different transitions, and thus could simply fall out of the trap (or be repelled by it, depending on the transition frequencies), a process called *photoassociation* [61]. Or, a collision between atoms could swap which ground state one of the atoms was in. In this process, known as *hyperfine changing collisions*, the energy difference between ground states is converted to kinetic energy [94]. For rubidium, the ground-state splitting is on the order of a few gigahertz (times Planck's constant). Dividing by Boltzmann's constant to convert to a temperature, we find this corresponds to the scale of a Kelvin to a tenth of a Kelvin, far larger than the millikelvin depth of our trap.

Those two processes tend to dominate when the atoms are exposed only to far off-resonant light, such as the trapping beam. Another pair of processes tend to dominate when the atoms are exposed to near-resonant light. One of these processes, called *fine structure changing collisions*, involves the atom being excited by the beam, and then changing to the other excited state by a collision, with some of the energy difference being converted to sufficient kinetic energy to eject the atom from the trap. In the other process, which is known as *radiative escape* and tends to be the dominant loss mechanism in conditions somewhat similar to our one-way barrier, the beam excites one atom, increasing the atom-atom attraction [61, 93]. Under this different interaction, the atoms accelerate towards one another. Should the atom decay back to the ground state, the atoms return to a more non-interacting state, but keep the extra kinetic energy, which is also (typically) sufficient to eject the atoms from the trap.

While we suspect that radiative escape is the dominant loss mechanism in our one-way barrier, we note that, for our purposes, it does not matter. All these loss rates tend to be proportional to the probability of exciting an atom while there is another atom nearby, and so they tend to be collectively labeled *light-assisted collisions*, or a subset of what are known as *density-dependent losses* (loss mechanisms where the probability of losing an atom depends on the local density of atoms). As such, the effects can be mitigated by choosing a one-way barrier mechanism that does not involve beams that can excite atoms as easily. Our experimental setup was designed to focus more on trapping and performing position measurements on small numbers of atoms, and is not particularly well set up for either a detailed investigation of different implementations of optical one-way barriers and their loss mechanisms.

Rather than attempt an in-depth exploration of the particular type of loss and ways of mitigating it, we simply satisfied ourselves that a generic density-dependent loss rate could explain our shorter lifetimes, as that strongly suggests a way to increase the lifetimes by avoiding methods that atomic excitation as common as it is in our setup. We satisfied ourselves of this by simulating a density-dependent loss rate that is proportional to the likelihood of exciting an atom in the trap, which is a generic approach that encompasses all of the loss mechanisms described above.

We implemented the light-assisted collisions in post-processing, after the simulation was run. One argument for this is that the we had the simulation running with very small time steps, while the light-assisted collision losses, as can be seen by the lifetime in Figure 4.19, happened on a much larger time-scale (but there is no reason we could not have had the simulation only do the light-assisted collisions every $N > 1$ time steps). The simulation output contains a record of the position and velocity of every atom in the simulation, but with only one update for a large number of time steps. With each update, it also recorded the number of time steps each atom spent in each state. This allowed us to go back in post-processing and calculate what fraction of the time each atom spent in the excited state, which is a function both of what fraction of the time the atoms spent in each state and where in each beam the atom was. We then estimated the local density of atoms by counting the number of atoms in several small regions near the beams, and used linear interpolation to get a density as a function of position. Using that, and the fraction each atom had spent in each state, along with the known profiles of the beams, we were able to integrate a loss probability for each output time step. With a final pass, we marked atoms in the simulation as having been ejected from the trap, using the computed ejection probability, and then simply skipped those atoms when counting the number of atoms on each side of the barrier.

Modeling light-assisted collisions using our method has only one free parameter. Once we have estimated the local density of atoms and what fraction of time the atoms spent in the excited state, there is a scaling parameter that multiplies by those factors to give the actual probability of a loss. This parameter could be thought of as the range of the interaction between atoms, and determines what it means for a density to be high enough to start causing losses. In our case, we fit this parameter to the experimental data shown in Figure 4.19. We then compared our value to one measured by the Wieman group [61]. The coefficient they reported, $K = 1.1\,(5) \times 10^{-10}$ cm^{5}mW^{-1}s^{-1}, includes the squared detuning (resulting from the probability of exciting the atom in the far-detuned limit), and so, once we include that factor into our parameter, we get a value of 2×10^{-9} cm^{5}mW^{-1}s^{-1}. The actual density of atoms in our simulation is low enough that we really cannot approximate the loss rate very well using our post-processing method, especially since the density of atoms in the simulation is different from that in the experiment. Given our rather crude method of modeling light-assisted collisions, we consider matching a measured value to within an order of magnitude to be quite good.

We will finish this section on simulations with a more thorough discussion of how angular momentum appears to affect atoms in our trap, pointing out some odd limits in which the atoms behave quite oddly. For this discussion, we will discuss only the single trapping beam, without the barrier beam and repumping beam, and ignoring the effects of gravity to preserve cylindrical symmetry (some of the amusing limiting cases are sensitive to gravity). We will assume the beam produces a perfectly Gaussian and conservative potential (with no scattering). This lets us write the potential (which is proportional to the intensity given by Equation (IV.3)) in cylindrical coordinates about the beam axis:

$$\begin{aligned} V(z,r) &= -V_0 \frac{e^{-q^2}}{1+\left(\frac{z}{z_0}\right)^2} \\ q^2 &:= \frac{2r^2}{w^2\left[1+\left(\frac{z}{z_0}\right)^2\right]}. \end{aligned} \tag{IV.6}$$

Here, z is the distance along the beam axis, with $z = 0$ being the trap focus, z_0 being the Rayleigh length of the trap, r is the radial distance from the beam axis, and w is the $1/e^2$ intensity radius of the trap at the focus. The abbreviation q represents the radial coordinate, scaled by the width of the trap at that z value, and will occur several times in the following discussion. The potential is assumed to be attractive, and so we state that $V_0 > 0$.

The equation of motions are simply derived from Equation (IV.6), and so we will simply show the result.[16] As in the orbital mechanics (or central potential) problems common in classical mechanics textbooks, the potential is independent of one of the coordinates (the azimuthal coordinate) [95, 96]. The equation of motion for that coordinate yields a conservation of angular momentum about the trap axis:

$$J := mr^2 \frac{\mathrm{d}}{\mathrm{d}t}\phi. \tag{IV.7}$$

where ϕ is the azimuthal coordinate, and m is the particle mass. The three equations of motion work out to be:

$$\frac{\mathrm{d}^2 z}{\mathrm{d}t^2} = -\frac{2V_0}{mz_0^2} z \left(1 - q^2\right) \frac{e^{-q^2}}{\left[1 + \left(\frac{z}{z_0}\right)^2\right]^2} \tag{IV.8}$$

$$\frac{\mathrm{d}^2 r}{\mathrm{d}t^2} = \frac{1}{m^2 w^4 r^3} \left[J^2 - mw^2 V_0 q^4 e^{-q^2}\right] \tag{IV.9}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} J = 0. \tag{IV.10}$$

Inspecting this, we can see that for small r, with no angular momentum, there is a linear restoring force in the radial direction (q^2 is proportional to r^2, and so the q^4 divided by the r^3 in the denominator leave a linear factor). As we might expect, though, for larger angular momentum, the radial equation allows for stable circular orbits about the trap axis (assuming z is held fixed, which it is not). The longitudinal equation also shows a linear restoring force, but that restoring force is reduced by the $(1 - q^2)$ factor. In fact, for large enough r, it actually becomes a repelling force. This is because, far enough from the axis, the beam intensity actually gets brighter away from the focal plane because the widening of the beam leaving the focal plane becomes a stronger effect than the fading due to the wider A particle with large r but no angular momentum will oscillate through the trap

[16] They are easy to derive using the Lagrangian formulation. Trying to derive them directly using Newtonian mechanics is trickier, but quite possible. This is because the direction of increasing radius (or azimuthal angle) is dependent on where a particle is, which is not true in Cartesian coordinates. As an example, consider a free particle moving at constant speed s parallel to the x-axis, with $y = 1$ (so $x = st$ and $y = 1$). The acceleration is zero, but this is not obvious from the cylindrical coordinate representation where the radius is equal to $\sqrt{1 + (st)^2}$. Using Newtonian mechanics, we must account for the shift to non-Cartesian coordinates by explicitly accounting for the change in directions as well as the change in coordinates. This conversion is built-in to the Lagrangian formulation (it could even be considered the motivation for their derivation).

axis, and so see a mix of attractive and repelling forces along the z-axis, but one with sufficient angular momentum will see a reduced attractive, and perhaps even a repelling force. This extra change in apparent longitudinal force accounts for the extra dephasing and longer period we see in the trap.

We finish with a quick observation. Our simulations can reproduce to fairly high accuracy a circular orbit about the trap axis that with a fixed, non-zero z value. We do this by noting that $q^2 = 1$ sets the longitudinal "force" (Equation (IV.8)) to zero, which we can do for arbitrary z by picking r appropriately. We can then pick a azimuthal velocity (which determines J) to force the radial "force" (Equation (IV.9)) to zero. This orbit, however, is unstable. The kinetic energy of the orbit is entirely in the azimuthal motion, which works out to be $J^2/2mr^2$. Since J is picked so that the term in brackets in Equation (IV.9) is zero, we can use the definition of q (and the fact that $q^2 = 1$ for this orbit) to solve for the kinetic energy. The result exactly cancels the potential energy, so that the total energy is zero. This means that the particle is technically not trapped. Furthermore, because there is a continuum of (r, z) pairs that satisfy $q^2 = 1$, and J is conserved, the equations allow for a particle with small deviations from that orbit to drift along that continuum. Our simulations show that the orbit is not particularly stable to deviations. If q^2 is not exactly unity, then this tends to be treated as a perturbation in r, which tends to have a stiffer radial force, and the r value oscillates rapidly about a value that does have $q^2 = 0$. This samples a range of restoring forces in the z direction, and tends to cause the particle to drift away from the focus. If J is perturbed to be slightly larger instead, then the total energy of the particle is positive, and the particle is not bound. In this case, it drifts away from the focus, traveling along the $q^2 = 0$ manifold, never to return. If J is slightly less (keeping a circular orbit, which means we break from $q^2 = 1$), the particle starts a slow oscillation about the focus of the trap. As a final killer preventing this particular orbit from being seen, because this orbit is rather unstable, we suspect that gravity would disturb the symmetry enough to make this orbit hard to see. Despite the unlikelihood of observing a repelling focus, several atoms in the simulations shown in Figure 4.17 were effected strongly enough by this effect that they never managed to cross through the focus, even with the 100 ms experiment time plus the initial half-period before the barrier beams were enabled. At least one atom in the simulation seemed to reflect off of the focus of the trap.

Accounting for the Entropy Changes in an Optical One-Way Barrier

Our one-way barrier is a simple modification of the classic Maxwell's Demon. James Maxwell, in his *Theory of Heat*, described a theoretical being (now referred to as Maxwell's Demon, although not called that in Maxwell's book) who knew the location of every molecule of gas in a divided container [97]. By operating a little trap door in the divider of the container, the demon could allow warmer molecules to pass from left to right, and colder molecules to pass from right to left, but not the other way around. In doing so, this demon could separate hot atoms from cold atoms. In theory, once that separation is made, the temperature difference could be used to run some sort of engine, and the process could be repeated *ad infinitum*, allowing one to extract thermal energy from something that is initially at a constant temperature.

Such a device is considered to be impossible by invoking the Second Law of Thermodynamics, which states that the entropy of a closed system cannot decrease. Thermal equilibrium is the highest-entropy state of a closed system, so the demon, without expending energy, is decreasing the entropy of the system by sorting the molecules into hot and cold, which is not thermal equilibrium. Our one-way barrier also decreases the entropy of the system, but by decreasing the volume occupied by the atoms rather than sorting by temperature. This decrease in entropy can also be used to run an engine off of a constant-temperature reservoir, as shown in Figure 4.20. We begin by inserting a one-way barrier into a container of atoms. Through its normal motion, each atom will eventually end up on the reflecting side of the barrier, where it will remain. Eventually, all the atoms will be trapped on the reflecting side. The force of atoms reflecting off of the barrier exerts a pressure on the barrier. Initially, with atoms on both sides, there was a pressure on both sides resulting in no net force, but there is now a strong force attempting to expand the container. In the case of our optical one-way barrier, we could then insert something solid for the atoms to push, but here we pretend the barrier itself is something physical. By letting the atoms press the barrier, we can extract work from the atoms in the container. The atoms, reflecting off a receding wall, reflect with less kinetic energy than they had, cooling the atoms. Thus, after expansion, the atoms are cooled and their thermal energy has been transferred into pushing the barrier. At this point, we have extracted work from a system that was initially at thermal equilibrium. Often, the explanation goes that the chamber is in some thermal bath, which heats the atoms back to their original temperature, and is then in the exact same state as at the beginning of the process. Either way, this shows a method of converting thermal energy into useful mechanical work in a system at a single temperature.

To show why such a system violates the second law of thermodynamics, we will give an explanation of what entropy is, and then use that to show why the above systems cannot work without either:

1. adding more energy to the system than the system produces, or

2. having access to something that is at a lower temperature than the system, in which case the work comes from heat flowing from a high temperature to a low temperature, which is not a violation of the second law.[17]

Interestingly, while Maxwell himself realized the demon he described in 1871 appeared to be a problem with the laws of thermodynamics, this problem was not resolved until the 1980s, when the presence of computers brought forth information theory [98, 99]. We will describe this resolution, and then then explicitly show how a system like our optical one-way barrier does not violate the second law of thermodynamics.

Entropy is often described as a measure of disorder; more formally, it is a measure of probability, likelihood, or missing information. When you have a system that

[17] Since the system does not really involve a transfer of energy, we do not really need a temperature imbalance as simply a way of transferring entropy that causes a larger increase in entropy in the reservoir than the corresponding decrease in the system. In our case, this involves a low-entropy reservoir where it is very easy to increase the entropy.

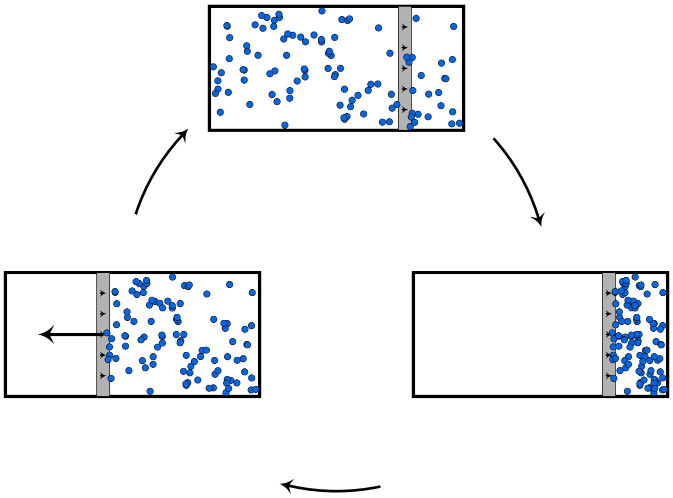


Figure 4.20. A simple apparently-perpetual-motion engine one could build using a one-way barrier. The one-way barrier divides the container, and eventually all atoms end up on one side. This creates a pressure difference, which could be used to operate a small piston. After expansion, the atoms we be cooler, having transferred kinetic energy into the piston. If the atoms could then be warmed up to some environmental temperature, the process could be repeated. This warming process injects the same amount of energy as was exerted on the piston, so energy is conserved. However, useful work is performed without increasing entropy, which essentially violates the second law of thermodynamics. We argue in the text that the entropy of the one-way barrier must somehow increase to compensate, and explicitly show how in the case of our one-way barrier.

could be in any of N states, with a probability $\mathcal{P}_i$ of being in a particular state i, then the entropy, S, is given as:

$$S = -\sum_{i=1}^{N} \mathcal{P}_i \ln(\mathcal{P}_i). \tag{IV.11}$$

This definition of entropy may be found in many textbooks, although it is sometimes treated as a derived formula [100]. Our treatment of entropy as it pertains to our barrier is our own.

One example to show how Equation (IV.11) is a measure of disorder, likelihood, or information, is to work with a particle in a box. In classical mechanics, a particle is completely described by its position and momentum at a given time (assuming no internal degrees of freedom). Given that, the available states for the particle are the set of all possible positions and momenta for the particle, as long as the position is inside the container (the one thing we know about the system). This is actually an uncountable set, so it is not obvious how to apply Equation (IV.11) to compute the entropy. The solution is to coarse-grain the system, and divide the physical space inside the container into N equally sized volumes. We can likewise divide the set of all momentum vectors into some three-dimensional grid, and compute the entropy using Equation (IV.11):

$$S = -\sum_{i=1}^{N}\sum_{p}\mathcal{P}_{i,p}\ln(\mathcal{P}_{i,p}).$$

where the i sum covers the different volumes, the p sum covers the different momentum values (a three-dimensional sum), and $\mathcal{P}_{i,p}$ is the probability of the particle being at position i (in subvolume i) with momentum p. Note that we could also consider having a sum over all space, instead of just subvolumes of the container, but since the probability of the particle occupying positions outside the container is zero (that is what we know about the system), we can always restrict the sum to a volume within the container.

Since we know nothing more about the particle other than some restriction on its location in space (it is in the container), we assume that the probability of the particle being in a particular one of the N subvolumes is independent of which subvolume, or what the momentum is.[18] The particle has been bouncing around in there a while, and probably spent some time in every corner of the container, so, without further information, we do not know where it is, or where it is more likely to be, so we must assign constant probabilities to everything.[19] As described

[18] This is a variant of what is known as the *ergodic hypothesis* in physics [100]. Once you reach the limits of your knowledge of the system, you assume the system is equally likely to be in any of the states allowed by what you know about the system. This assumption can typically be justified by assuming the distribution of initial conditions is finely spread among the possible states, or that the evolution of the system allows it to sample enough of the available states that the probability of being in any of them (or near any of them) is roughly constant.

[19] We could imagine a container with a large open volume and a very small, long tube that twists and turns in some fantastically complex manner, such that the probability of a particle successfully navigating the tube are much less than the odds of defeating a Star Destroyer in the Millennium Falcon. In such a case, if we know the particle started in the main region of the container, we

in Appendix B, independent probabilities multiply ($\mathcal{P}_{i,p} = \mathcal{P}_i\mathcal{P}_p$), so the entropy reduces to:

$$
\begin{aligned}
S &= -\sum_{i=1}^{N}\sum_{p}\mathcal{P}_i\mathcal{P}_p\ln(\mathcal{P}_i\mathcal{P}_p)\\
S &= -\sum_{i=1}^{N}\sum_{p}\mathcal{P}_i\mathcal{P}_p\left(\ln(\mathcal{P}_i)+\ln(\mathcal{P}_p)\right)\\
S &= -\sum_{i=1}^{N}\sum_{p}\mathcal{P}_i\mathcal{P}_p\ln(\mathcal{P}_i)-\sum_{i=1}^{N}\sum_{p}\mathcal{P}_i\mathcal{P}_p\ln(\mathcal{P}_p)\\
S &= -\sum_{i=1}^{N}\mathcal{P}_i\ln(\mathcal{P}_i)\sum_{p}\mathcal{P}_p-\sum_{i=1}^{N}\mathcal{P}_i\sum_{p}\mathcal{P}_p\ln(\mathcal{P}_p)\\
S &= -\sum_{i=1}^{N}\mathcal{P}_i\ln(\mathcal{P}_i)-\sum_{p}\mathcal{P}_p\ln(\mathcal{P}_p).
\end{aligned}
$$

where we have used the fact that the sum of all the probabilities for a given distribution is unity. This result shows that, for our independent probability distributions over position and momentum, the entropy breaks up into the sum of a position-only entropy (using a sum over the subvolumes and the probabilities of being in those) and a momentum-only entropy. Now, if we let $\mathcal{P}_i = 1/N$, since all N subvolumes are equally likely, the first term reduces further:

$$
\begin{aligned}
S = -\sum_{i=1}^{N}\frac{1}{N}\ln\left(\frac{1}{N}\right)-\sum_{p}\mathcal{P}_p\ln(\mathcal{P}_p) &= \ln(N)\sum_{i=1}^{N}\frac{1}{N}-\sum_{p}\mathcal{P}_p\ln(\mathcal{P}_p)\\
&= \ln(N)-\sum_{p}\mathcal{P}_p\ln(\mathcal{P}_p). \qquad \text{(IV.12)}
\end{aligned}
$$

So far, our entropy of a particle-in-a-box, Equation (IV.12), has two essential problems, solved by a common solution. First, the volume term $\ln(N)$ is dependent on the size or number of subvolumes into which we split the container, and approaches infinity as the number of subvolumes tends to infinity. Second, no matter

could assign a lesser probability to the particle being in the tip of the tube than to an equivalent volume in the main part, since the probability of the particle making its way down the tube is so small. However, if we do not know anything about the starting conditions, then it is possible the particle started in the tube, or at least started so long ago that it could have made its way down the tube. In this case, we must assign equal probabilities.

how coarsely we split the available momentum values, the number of momentum "subvolumes" is infinite, and so the equivalent $\ln(N)$ term is already infinite. One possible solution is to invoke a little more knowledge about the system. For example, the energy is not infinite, so it is reasonable to expect that the momentum is, in fact, bounded.[20] Also, if we look at combined position and momentum states in phase space, as opposed to separating them as we did here, quantum mechanics and the Heisenberg uncertainty relation suggest a natural phase-space volume to use. These two effects give a finite entropy, but if they seem rather arbitrary, that is fine. Entropy is most useful in comparing systems (or states of the same system). If the momenta distributions of two systems are the same, then *any* method we use to make the momentum term of the entropy finite, as long as we are consistent, will make that term of the entropy of each system the same. Thus, in comparing entropies, if the momenta distributions are the same, those terms cancel. If the momenta distributions are different, then we care only about the *difference* in entropies of the resulting entropies, and the problems with infinity tend to disappear.

So, with that in mind, we will define some arbitrary subvolume of the container with size v, and some arbitrary subvolume of momentum space with size p, with some total volume V_p. As used above, N is the number of subvolumes in the container, which now becomes V/v. Likewise, the same derivation of $\ln(N)$ for the position space works on the momentum term, and we arrive at this version of entropy for a particle in a box:

$$S = \ln(V/v) + \ln(V_p/p) = \ln(VV_p) - \ln(vp).$$

With one final modification, we will have derived the entropy of a particle in a box. The term $\ln(vp)$ depends on our arbitrary choice in subvolume size, which we will presumably use everywhere we compute entropy. Thus, it is a constant term which will be applied to every entropy, and we can therefore drop it. Our final entropy for a particle in a box is therefore:

$$S = \ln(\Omega). \tag{IV.13}$$

[20] Another assumption is to drop the strict ergodic hypothesis and assume that the probability of having really high momenta goes to zero smoothly as the magnitude of the momentum increases.

Here, we have replaced the product VV_p by Ω. The product VV_p resulted from our earlier decision to handle positions and momentums separately. Had we treated the problem with those degrees of freedom combined, we would have found ourselves splitting the entire available phase space into subvolumes, with equal probability distributed among them, and ended up with some entropy equal to the logarithm of the available phase space volume (minus the logarithm of some constant subvolume in phase space, which we would have dropped). This is a slightly more general version of the product of volumes given here.

The derivation of Equation (IV.13) can be generalized to any system of non-interacting particles in some bounded region of phase space (the combination of all possible positions and momenta of every particle). The entropy of the entire system is the logarithm of the volume of the available phase space, where the volume includes the positions and momenta of every particle in the system. Equation (IV.13) is how entropy is often defined in thermodynamics. For any system with some number (or continuum) of states each with a constant probability, our original definition of entropy (Equation (IV.11)) reduces to the logarithm of the number (or volume) of states available, plus or minus some constant. If state A contains more substates or phase space volume than state B, then the entropy associated with state A is larger than the entropy associated with state B. If we have no other reason to assume state A is more or less likely than state B, we could assume the underlying substates were all equally likely, in which case the state with more substates is the most likely. In other words, the state with the highest entropy is the most likely.

For example, let us pick two states. State A is where the particle is in the left half of the box, and state B is where the particle is in the left-most quarter of the box. Knowing only that the particle is somewhere in the box, we do not know the particle is in either state A or state B. However, the probability that the particle is in the left half is larger than the probability that the particle is in some smaller region, so we state that state A is more likely than state B. This is the sense that entropy is a measure of likelihood, or probability. State A, having a larger volume, also has a larger entropy, since, in this case, entropy is essentially the logarithm of the volume. The most likely state is the one with the highest entropy.[21] This is why Equation (IV.13) is the definition of entropy often supplied in textbooks, as it is simpler to use, and makes the connection between entropy and likelihood more transparent.[22] We note, though, that when a system has a set of states where some states are more likely than others, this definition fails. Sometimes we can cheat and define the "volume" of each state as some measure proportional to the likelihood of each state, but it is often safer to simply resort to using the original definition, Equation (IV.11), which we will do shortly.

We are now prepared to describe why we also refer to entropy as *missing information.* We start with a set of states describing a system. Any one state completely describes the system; if we know the system is in one particular state, we consider the system fully described. We then compute the entropy based on how many states the system could be in (and the probability of being in each state). The resulting value is something like the logarithm of the number of states the system could be in. If we know exactly which state the system is in, the entropy is $\ln(1) = 0$. The more states the system *could* be in, the larger the entropy is. In this sense, a larger entropy means we know less about the system, or that there is more we could learn about a system. A system with an entropy of $\ln(2)$ essentially has two states it could be in. A system with the much larger entropy of $\ln(1000000)$ has a million states it could be in. The latter system, with the larger entropy, is much less determined than the former, which means there is much more we could learn from it than the

[21] Note that this analogy is not perfect without some qualifications. For example, say state A was the particle being somewhere within the box, and state B is the particle being in some larger volume that includes the box. Since we know the particle is in the box, both states are equally likely with probability 1, even though the entropy for state B is larger since it refers to a larger volume. Technically speaking, this example is invalid. If we know the particle is in the box, then not all the volume covered by state B is equally likely to be occupied, so we cannot use $S = \ln(\Omega)$ for computing entropy. Done correctly, we would find both states have the same entropy.

[22] As we will briefly cover, entropy in thermodynamics is often associated with energy and temperature, so, to aid in conversion, entropy is often defined as $S = k_\mathrm{B}\ln(\Omega)$ instead of just $\ln(\Omega)$, where k_B is Boltzmann's constant. In this form, ST is the energy required to transfer an amount of entropy S into a reservoir at constant temperature T. However, since this section mostly deals with entropy as a measure of information, we choose not to include k_B in the definition. Texts sometimes skirt the issue by measuring temperature in units of Boltzmann's constant; in such units, $k_\mathrm{B} = 1$.

first, where we can only learn one more piece of information (which state it is in).

This is also why entropy is considered a measure of disorder, although this is a poor measure of entropy as the concept of disorder can be non-intuitive. As an example, we consider an ice cube. In our overly simplistic model, each molecule of water in the ice cube is perfectly placed in a crystal lattice. Once the orientation and location of the cube is set, the locations of all the molecules are set. If we knew exactly where all the molecules are, then the entropy of the ice cube is zero. Now, if we drop the ice cube and it breaks all over the floor, we have many smaller ice cubes. In order to specify the location of the all the molecules, we need to describe the location, orientations, and shapes of all the pieces of ice. Now, there is much more information that needs to be given to determine the location of all the water molecules. If we specify all that, the entropy is once again zero, but if we simply know that the cube broke into many pieces, there are many ways the cube could have broken, and so the entropy is now much larger. The more pieces of ice there are, the larger the entropy. If we knew the ice pieces would all be the same shape and size, and would align themselves on some grid pattern, that is a very small subset of all the possible ways the pieces could end up. Worded another way, the number of ways the ice could break that appear "ordered" to the human mind is a very small subset of the total number of ways the ice could break. Therefore, the entropy of an "ordered" breakage is much lower than a "random" breakage. This concept breaks down for most people in the limit where the number of pieces is the number of molecules. This is essentially water, which has a much higher entropy than broken ice cubes, since we now need to specify individual locations for every molecule as opposed to clusters of them. However, since this "disorder" is well below what we humans can detect, we mostly see water with a smooth surface, which looks "ordered" even though it has a much higher entropy.

With these descriptions of entropy, the second law of thermodynamics is almost self-evident. Say we start with two boxes of the same size, one empty and the other with a known number number of particles with a certain total kinetic energy. If we connect the two boxes so that particles may travel between the two, eventually the particles will spread out to fill the two boxes. Assuming that each particle occupies one of the boxes with probability 1/2, independent of any other particle, we can compute the probabilities of the particles being in one box or the other. These probabilities are plotted in Figure 4.21 for varying numbers of particles. Note how sharply peaked the distribution becomes about the case where the particles are nearly evenly split. For a small number of particles, it is quite possible that all the particles will be in one box or the other (for $N = 1$, it is a certainty). For a large number of particles, though, it is effectively an absolute certainty that the particles will be split approximately evenly between the two boxes. This is a binomial distribution, which, as described in Section B.5, has a standard deviation of $\sqrt{N}/2$ (for a probability of 1/2), and is nearly Gaussian for large N. For a Gaussian, approximately 99.7% of the distribution is contained within 3 standard deviations of the mean.[23] Therefore, with better than 99% certainty, we can expect half the particles, plus or minus $3\sqrt{N}/2$, to be in one box. For a thousand particles, that deviation is about 47, or about 5%. For one million particles, that deviation is less than 0.2%.

The second law of thermodynamics states that entropy of a closed system never decreases. Technically, this is not true. In our two-box example, the highest entropy case is where the number of particles is exactly split between the two boxes (or as close as is possible for an odd number of particles). All the time, particles can change side. Unless pairs always swap sides together, there will be deviations from that perfect state, and those changes must decrease the entropy. However, for a large number of particles, the probability of a substantial deviation from the maximum entropy state is so tiny as to be ignorable. Considering that the initial state was one where all the particles were in one box, which we are calling an entropy of 0, it is fair to say that, in this case, the entropy has risen and will never decrease (noticeably) again without some external influence.

The second law presents a distinction we will use later. A *reversible* process must not increase entropy. If a process can just as easily go one direction as another,

[23] The integral of the Gaussian, called the *error function*, is well-tabulated and most numerical software has some way to compute it. However, it is handy to remember that approximately 68% of the area of a Gaussian is within one standard deviation of the mean, and 95% and 99.7% are within 2 and 3 standard deviations, respectively.

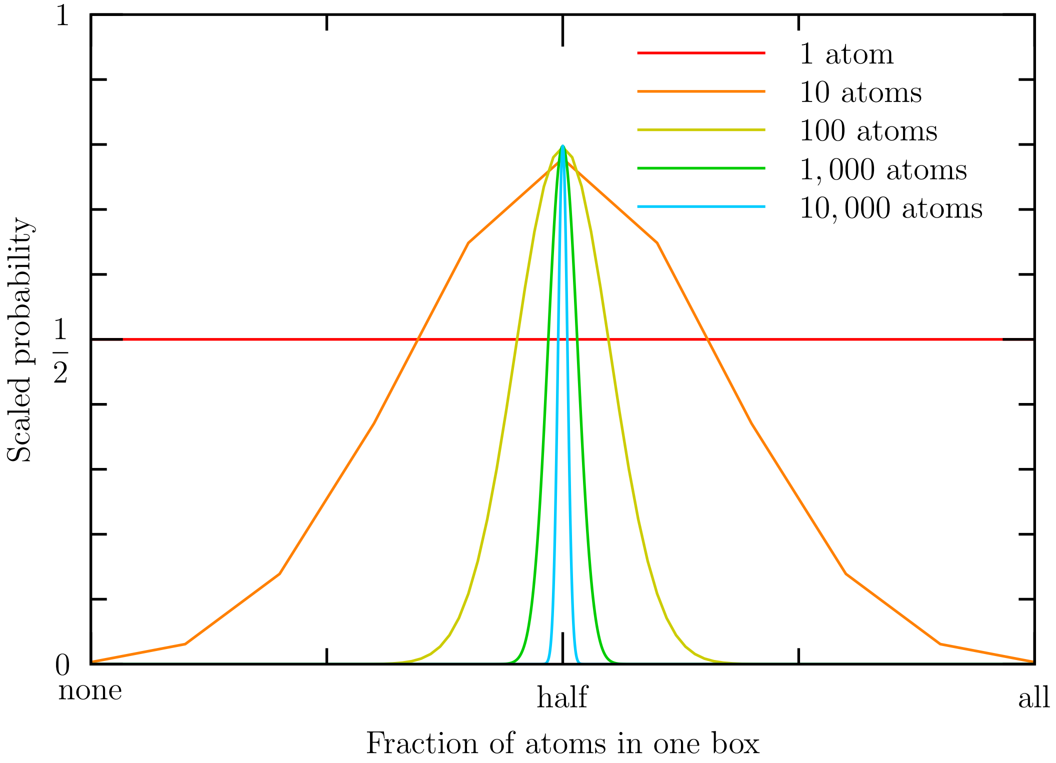


Figure 4.21. The probabilities of particles being in one of two boxes. We plot the probability of having a certain fraction of N particles in one of two equally likely connected boxes, for various value of N. For a single particle, the probability is 1/2 that the particle is in the box (all the particles are in the box) and 1/2 the particle is not in the box (none of the particles are in the box). For larger values of N, the distribution becomes very sharply peaked about 1/2, where the particles are split evenly between the two boxes. The horizontal scales have been adjusted so that all are on a scale from "none" to "all" particles being in the box. The vertical scales were adjusted according to the approximate width of the distribution to compensate.

then that process must not increase entropy, for it would need to decrease entropy to reverse itself. A ball rolling down a hill in an ideal frictionless environment is a reversible situation. We could just as easily imagine the ball bouncing off a lossless spring at the bottom and then rolling back up the hill. This is a reversible process, and so must not involve an increase in entropy. An *irreversible* process, on the other hand, must increase entropy. Otherwise, the final collection of states is just as likely as the initial. This means that, at some point in the future, the system is just as likely to be back in the initial state, and so the process was, in fact, reversible. Therefore, for a process to be irreversible, the final state must have a higher entropy than the initial.

We will now quickly demonstrate some uses of entropy in preparation of our calculation of the one-way barrier entropy. If we have two systems A and B and allow energy to transfer between them (but not in or out of the pair of systems), then the second law states that energy will flow between the two systems until the total entropy is maximized, in analogy with our particle case given above. We represent the total energy as E, and the energy in each system E_A and E_B, with $E_A + E_B = E$. Likewise, the entropies of the two systems are $S_A(E_A)$ and $S_B(E_B)$. The entropies could easily be functions of other variables, such as number of particles, so we will use partial derivatives in our definitions, but only explicitly show energy dependence. The main caveats are that E is constant, and the systems are distinguishable, allowing us to say S_A is *not* a function of E_B, and vice versa.

We now ask what the equilibrium energy of each system is. The two system case is the simplest. We can use the conservation of energy (E is constant) to eliminate all but one variable, so that we are trying to maximize:

$$S_A(E_A) + S_B(E_B) = S_A(E_A) + S_B(E - E_A).$$

Assuming the functions are differentiable, the maximum occurs when the derivative of this quantity with respect to E_A is zero.[24] Thus, we arrive at this relation for

[24] The other possibility is that the maximum occurs at some boundary value of E_A, such as 0 or E if we are disallowing negative values of energy. These endpoints require that the energy of one of the systems reaches the lowest possible value for that system. In practice, entropy functions

the final state of the two systems:

$$\frac{\partial}{\partial E_A} S_A(E_A) + \frac{\partial}{\partial E_A} S_B(E - E_A) = 0.$$

Using the chain rule to substitute back to E_B, the final result is:

$$\frac{\partial}{\partial E_A} S_A(E_A) = \frac{\partial}{\partial E_B} S_B(E_B).$$

In fact, for the general case of multiple systems, the equilibrium solution is where this relation is satisfied:

$$\frac{\partial}{\partial E_A} S_A(E_A) = \frac{\partial}{\partial E_B} S_B(E_B) = \frac{\partial}{\partial E_C} S_C(E_C) = \cdots .$$

This is because if the derivatives are not equal, we can always increase the total energy by shifting a small amount of energy from a system with a smaller derivative to a system with a larger derivative.

When we connect these systems, energy tends to flow from the systems with the smaller derivatives to the systems with the larger derivatives, until all the derivatives are equal, simply because there are so many more states where that is approximately true than not. This is what we observe to be true about temperature (albeit in the wrong direction). In thermodynamics, temperature is often *defined* in terms of these entropy derivatives. However, the usual definition is that energy flows from high temperature to low temperature, which is backwards from the derivatives, so the usual definition for the temperature T of a system flips that:

$$\frac{1}{T} := \frac{\partial}{\partial E} S(E). \tag{IV.14}$$

In other words, the temperature of a system is reciprocal of the partial derivative of the entropy as a function of energy. If we can invert $S(E)$ and write the energy of a system as a function of the entropy ($E(S)$), then by the chain rule, the temperature is also:

$$T = \frac{\partial}{\partial S} E(S).$$

typically have infinite positive slopes at the lowest possible energy. This is the essence of the third law of thermodynamics, which is outside the scope of this discussion. What it means here is that these boundary values are not maxima, as the entropy increases infinitely fast when a little energy is added to the system and more than compensates for the corresponding decrease in entropy for the systems from which the energy comes.

Note that, as we have defined entropy to be unit-less, temperature has units of energy. This is not how we usually define temperature, and so a scale factor called Boltzmann's constant is often introduced as a conversion between energy and temperature. This factor is often included in the definition of entropy, which makes Equation (IV.14) correct. The definition of temperature with an explicit Boltzmann's constant is

$$\frac{1}{k_{\mathrm{B}}T} := \frac{\partial}{\partial E}S(E),$$

where we use the unit-less entropy we have been using.

We can quickly show that these definitions are related to the physical world. One typical case is there is a single system of interest connected to a much larger system, called a thermal bath, or a *reservoir*. The important feature of the reservoir is that it is so much larger than the system that even if we dump all the energy of the system into the reservoir, the reservoir will not be greatly affected. As a simple example, we could take a single molecule of water (the system of interest) in a swimming pool (the reservoir). Clearly, to a very good approximation, transferring energy from the molecule of water to the rest of the pool and back, is not going to noticeably change the total energy of the pool. We assume the total energy of the system is a constant E_T. The system of interest (the water molecule in our example) has energy E and entropy $S(E)$. Because the reservoir (the pool in our example) is so much larger than the system, we assume the energy and entropy of the reservoir change little as energy transfers between the reservoir and the system, so we can write the system entropy as a first-order expansion in E:

$$S_R(E_T - E) \approx S_R(E_T) - \frac{\partial S_R}{\partial E}E.$$

The derivative is just the reciprocal of the temperature of the reservoir. By not including higher-order terms in our expansion, we are essentially assuming that the temperature is roughly constant as energy transfers back and forth between the system and reservoir, as stated by the second law of thermodynamics. Since the fluctuations about equal temperature get smaller as the system gets larger, this is particularly true for the reservoir, but not necessarily the system of interest. Making this identification, we have:[25]

$$S_R(E_T - E) \approx S_R(E_T) - \frac{1}{k_\mathrm{B} T_R} E.$$

Assuming the various states of the system of interest combined with the reservoir are equally likely, the likelihood that the system has energy E_1 is proportional to the number of states in which the system has energy E_1, a phase-space volume we will call Ω_1. We can use this along with Equation (IV.13) to evaluate the relative likelihoods of the system having two different energies E_1 and E_2:

$$\begin{aligned}\frac{\mathcal{P}_{E_1}}{\mathcal{P}_{E_2}} &= \frac{\Omega_1}{\Omega_2} = \frac{e^{S_1}}{e^{S_2}} \\ &= \exp(S_1 - S_2).\end{aligned}$$

We now apply our approximation that the reservoir is much larger than the system of interest, allowing us to use our above approximation for the reservoir entropy and to ignore the entropy contribution of the system of interest. The constant entropies cancel, leaving us with:

$$\frac{\mathcal{P}_{E_1}}{\mathcal{P}_{E_2}} = \frac{e^{-E_1/k_\mathrm{B} T_R}}{e^{-E_2/k_\mathrm{B} T_R}}.$$

Our approximations get better as the reservoir gets larger (assuming we can wait long enough for it to reach thermal equilibrium). It therefore makes sense to assign relative probabilities to a system of interest in a much larger thermal bath having a certain energy E:

$$\mathcal{P}_E \propto e^{-E/k_\mathrm{B} T}. \tag{IV.15}$$

Here, we have derived the famous Boltzmann factor. In our example of a molecule in a swimming pool, it can be used to give the kinetic energy distribution of the molecule. In a chemical reaction, it gives the relative proportions of the various chemical products; in the atmosphere, it gives a decent approximation of the density of air with altitude, although that is an approximation because the atmosphere is

[25] Note that we can drop the subscripts and write $T k_\mathrm{B} \Delta S = -\Delta E$. As mentioned, Boltzmann's constant is typically included in the entropy, and we see that in these constant-temperature cases, $\Delta S T$ is the amount of energy we must transfer into the constant-temperature reservoir to increase its entropy by an amount ΔS.

not at thermal equilibrium and is not a closed system as sunlight heats it and it radiates energy into space. In the altitude example, the potential energy of a particle with mass m in a constant gravitation acceleration g at a height h is mgh, and so, in a limited region of the atmosphere, the probability of a molecule of air being at height h would scale as $\exp(-mgh/k_{\mathrm{B}}T)$. Since air molecules are relatively non-interacting, this gives the height-dependence of the density of air.

Our atmospheric density example fails for our swimming pool example because the particles in a liquid are interacting particles. For water, the probability of a molecule of water being at the bottom of the pool is *not* independent of the positions of other molecules, since if the bottom of the pool if filled, our molecule cannot occupy that space.[26] We can, however, assume the velocity is rather independent of other particles (this is even more true in a gas). In that case, we can average over the possible kinetic energies of the particle, either in a pool or the atmosphere, where the probability of having that kinetic energy is $\exp(-E/k_{\mathrm{B}}T)$. If we assume all kinetic energies have similar phase spaces, then the average kinetic energy is $k_{\mathrm{B}}T$. More precisely, though, the equally likely states typically correspond to individual velocities. In three-dimensions, a particular energy corresponds to the surface of a three-dimensional sphere of velocities, with radius proportional to $\sqrt{E}$. The total sphere volume is therefore proportional to $E^{3/2}$, meaning the volume of a thin shell around that (the part having energy E) is proportional to $E^{1/2}\,\mathrm{d}E$. Thus, the probability of having a kinetic energy E is proportional to that volume times the Boltzmann factor, $E^{1/2}\exp(-E/k_{\mathrm{B}}T)\,\mathrm{d}E$. Taking the average kinetic energy with that yields $3k_{\mathrm{B}}T/2$, the average motional (we did not include rotational energy) kinetic energy of a molecule in a gas at temperature T. While the energy-volume prefactor differs from system to system, because the Boltzmann factor scales energy with the temperature, the end result will nearly always be that the average energy of the system is in some way proportional to the temperature. Thus, this odd definition of temperature in terms of entropy coincides with our usual intuition that temperature corresponds to the amount of heat (or energy) in a system.

[26] Back in air, there is enough of a gap between molecules that there is still plenty of room for another particle.

Now we see that a substantial decrease in entropy is, almost by definition, an extremely unlikely event. Therefore, the entropy of a system cannot decrease without the entropy of some other system increasing by a larger amount, for a net increase.[27] We can use this to describe why the engine described in Figure 4.20 cannot be used as a perpetual-motion machine. There is a system (the atoms in the container), and we use a one-way barrier to decrease the entropy of the system (or, in the case presented by Maxwell's demon, the decrease in entropy is a separation of hot and cold). That entropy change is used to produce useful energy. The system then re-absorbs that energy back from the reservoir, in order to repeat.

In the entropy-decreasing step, in order for that process to be spontaneous, the entropy of some other system (which we will call the memory, for reasons we will discuss shortly) must increase by an amount at least as large as the decrease. Then, in order for re-absorbing energy from the reservoir to occur, the entropy of the reservoir must decrease less than the increase in entropy of the system. *If* we ever need to return the memory to its original state, and we will argue that we will need to do just that, then we need to decrease the entropy of the memory. One option is to add more steps, but eventually, we will need to return that entropy to the reservoir. So, in summary, the system entropy decreases (we will call the change $-S$), but the memory entropy increases more. Restoring the memory requires an even larger increase in the reservoir entropy (which is true regardless of how many extra steps we need to insert), so the reservoir entropy increases by more than S. Restoring the system requires an increase of the system entropy of S, which means the reservoir entropy must decrease, but the change must be larger than $-S$. The important point is that the decrease in reservoir entropy is more than $-S$, and the increase is more than S, so the net change is positive. The system and memory are assumed to be restored to their original states, for a net entropy change of zero.

Since the reservoir is at constant temperature, we can use the definition of temperature in Equation (IV.14) to state that the energy transferred to/from the reservoir is the product of the temperature multiplied by the entropy change. This is because a constant temperature implies the derivative of energy with respect

[27] This is not guaranteed, but a safe bet.

to entropy is constant, so integrating that yields the result that energy change is proportional to entropy change. Therefore, the net energy transferred into the reservoir is the temperature times he net entropy change of the reservoir, which we just argued needs to be positive. This means that the energy the system can produce must be less than the energy we need to dump into the reservoir to get the system to work. In other words, the perpetual energy scheme fails, as it takes more work to operate the scheme than it produces.

We now discuss what we call the memory. In our case, that memory is the pumping beam, but in the more general description of Maxwell's demon, that is usually described as the memory of the demon, which needs to know the location and other properties of the atoms in order to separate them. Note that this memory must initially be in a known state (and so has a very low entropy), so that recording the position of atoms into it changes its entropy to one similar to the entropy of the atoms (which is a higher entropy), which is an increase of entropy. Otherwise, the process of writing to the memory must be irreversible. For example, any process that records a value of 1 into a section of memory with an unknown value (say either 0 or 1, for simplicity) is, by definition, irreversible. After all, once the 1 is written, if we were to try to reverse that write, there are two possible states (the 0 or 1 that the write process overwrote). As we cannot know which state the system was in based on the outcome, the process must have been irreversible, with an associated increase in entropy. This is equivalent to first erasing the memory and then writing to it, so we will just assume the memory must be erased first.

As long as we have erased "memory" to write into, we can continue the heat-engine cycle in Figure 4.20, and extract energy from our constant temperature reservoir.[28] However, if the memory is finite, we must eventually erase that memory and return it to a low-entropy state to start over again. That requires an entropy increase somewhere, and the only place left is the reservoir. If we can only transfer entropy to the reservoir through an energy transfer, then that requires the amount of energy described earlier. Note that we can bypass this energy requirement and

[28] Note, though, that by the following argument, it probably took an infinite amount of energy to create that erased memory in the first place.

achieve perpetual motion in two ways. We can either have a prepared supply of erased memory (which has an important energy cost), or we can have some way of transferring entropy to the reservoir that is not through energy (particle transfer, for example). Either way, we can only use the memory for "perpetual" motion until some supply of either memory, energy, or whatever vessel is used to transfer entropy to the reservoir, is exhausted. At that point, the "free" energy is done, and we are back to requiring energy (or something) to continue to get the process to work. Since the coupling to the reservoir at some point is usually energy, we state that a system like this can only work if we provide more energy that it produces, or we have a low-temperature reservoir we can dump into (essentially pumping heat from a high-temperature reservoir to a low-temperature reservoir), or we have an infinite memory (which is basically a low-entropy source).

We are now ready to show a computation of the entropy of atoms in our one-way barrier, and demonstrate that the second law of thermodynamics is upheld. In doing so, we will describe how we cannot extract work from such a system without a low-temperature source. In order to make the math easier, we make several assumptions. Each assumption either decreases the change in total entropy we compute, or we can show it retains a good approximation of the entropy. We then show that this decreased change of entropy is positive; therefore, the actual change in entropy is a net increase as well.

There are two components to our barrier: the atoms and the light fields used to trap them. The laser beams scatter light off the atoms, giving random momentum kicks which heat the atoms (increasing their entropy) and increase the entropy of the laser beam (which changes from a single beam to a spread). Even when acting as a conservative potential (which cannot decrease the entropy of the atoms), the atoms shift the phase of the laser beam. Since the ideal laser beam has a well-defined phase, any shifts from the atom should increase the entropy of the laser beams. The actual trapping potentials are, ideally, conservative, and could in theory be made arbitrarily close to that. The pumping beam that pumps from one state to the other requires scattering; however, with the use of cavities, we could theorize that it is possible to reduce the effects of the random kicks of scattering. These effects are not necessary to the operation of the barrier, and so we will ignore them. By showing that the entropy increases without these effects, we will automatically be showing that the entropy increases with them.

The one essential element of the atom–light interaction is that the pumping beam changes the internal states of the atoms. Our barrier cannot work without somehow changing the atoms; otherwise, the barrier beam cannot present a different potential to the untrapped atoms as the trapped atoms. Therefore, while we could theorize ways of removing the other effects of the interactions (and, in doing so, reduce the overall change in entropy), we must keep this one element. Furthermore, we need this transition to be *irreversible*. The pumping beam must be able to change untrapped atoms to the trapped state, but not the reverse (or, at least, the pumping beam must trap atoms more easily than it untraps them). Otherwise, the net effect of the barrier would not be to trap atoms in a smaller volume, and so would not really be a one-way barrier. Since the atoms end up in a smaller volume than they started with, we would expect the entropy of the atoms themselves to decrease. The only way the second law could hold is if some other component of the system had a larger increase in entropy. Because this change in atomic state is irreversible, it must be associated with some increase in entropy. This change is required for the operation of the one-way barrier, and so cannot be ignored. We will show that this process results in a net increase of the entropy of the pumping beam, and that that increase exceeds the decrease in entropy of the atoms, resulting in a net increase in entropy as predicted by the second law of thermodynamics.

We will assume that we initially have N atoms confined to a box of equal potential and volume V, so that each atom is equally likely to be in any given section of the box. Once trapped, the atoms will be confined to a fraction r of this box, where r is a real number between 0 and 1 (the trapping volume should be smaller than the initial volume). Furthermore, we allow that not all atoms will be trapped, and designate f as the fraction of the total number of atoms N that are trapped in the final state. Naturally, f is also bounded by 0 and 1, as we cannot trap a negative number of atoms, nor can we trap more atoms than there are available. Technically, f is limited to rational numbers such that fN is an integer, but our argument will not require that restriction. Certainly, if our expression for the change in entropy increases for all f, then it also increases for that subset which make fN an integer. Table 4.3 summarizes these variables.

In the case where the potential is non-uniform, and atoms are more likely to be found in certain regions, then we could think of this in an alternative space in which regions where atoms are more likely to be found are assigned more "volume" than they actually have, to make the probabilities work out right. We could also pull a similar trick to account for region-dependent velocity distributions. Regions of higher potential will have particles with less kinetic energy, so we would assign smaller "volumes" to those regions than regions with lower kinetic energies. The main point is that while we could create a more accurate model than our equal-potential model, the only use of volume in our model is some measure of likelihood of finding atoms there, so we only need to interpret V as proportional to the probability that an atom is in the entire trap (1), and rV as proportional to the probability that an atom is in the trapping region (r).

All of the atoms in our trap are assumed to be non-interacting. For cold rubidium at the densities in our trap, atom–atom interactions are very rare occurrences (see the discussion of that in Section IV.5), so this should be a good approximation.

We are considering the kinetic energy of the atoms to be constant throughout this process (or at least to occupy a fixed volume in momentum space), and so the only factors that affect the entropy of the atoms is the confinement volume, and the number of atoms. For a single atom, we can use our particle-in-a-box example in Equation (IV.13). The volume distribution is uniform and assumed to be independent of other factors, such as the kinetic energy distribution, so our single-particle entropy is:

$$S_{\mathrm{a}} = S_0 + \ln(V).$$

where we lump all the non-volume-related terms to the entropy in S_0.

We also assume that all the atoms are non-interacting and independent of one another, such that the entropy (and all distributions) of each atom is the same as for any other in the same state (trapped or untrapped). If we have 3 states for one

Variable	Range	Description
N	Whole numbers	Number of atoms, assumed to be large
V	Positive real numbers	Volume in which atoms are confined
f	$0 \leq f \leq 1$	Fraction of atoms that get trapped
r	$0 \leq r \leq 1$	Fraction of volume in which atoms are trapped

Table 4.3. A summary of the variables used to describe the entropy of the atoms in the one-way barrier.

atom, and 3 states for another, the number of states for n atoms is 3^n. Adding multiple particles exponentiates the number of states. Since the entropy is the logarithm of the number of states, our entropy for N non-interacting atoms is

$$S_{\mathrm{a}} = NS_0 + N\ln(V),$$

which is just N times the entropy for a single atom.

Finally, we throw in a term to account for the quantum mechanical indistinguishability of atoms.[29] Multi-particle states in quantum mechanics are typically symmetrized in such a way that swapping any particles does not change the state. For N atoms, there are $N!$ ways to reorder the atoms, and we count all of those as a single state.[30] Reducing the number of states by $N!$ means we subtract the logarithm of that from the entropy. Ordinarily, we would simplify by applying Stirling's

[29] We might also choose to treat all the atoms as distinguishable. We will point out later that the distinguishable case leads to a larger entropy increase, and so, since this case turns out to be a non-negative change, the distinguishable case must as well.

[30] We do not need to invoke quantum mechanics to get these terms. Shortly, we will have atoms in either the trapped or untrapped states, two different states, and we will be concerned simply with how many atoms are in one state or the other. There is only one configuration where all the atoms are in a particular state, but if we have fN in one state, and $N - fN$ in the other, then the total number of states within this collection of states is $\binom{N}{fN}$. Since we assume each of these states is equally likely, the effect on the entropy is to just add in the logarithm of the number of states, which works out to be $\ln(N!) - \ln(fN!) - \ln(N - fN!)$. The first term is constant and will eventually subtract out; the remaining two terms are exactly the terms we get from assuming the

approximation:

$$N! \approx \sqrt{2\pi N}\left(\frac{N}{e}\right)^N. \tag{IV.16}$$

This approximation is surprisingly accurate for larger N, with the error being a multiplying factor of approximately $\exp(1/12N)$. Thus, to a very good approximation, the logarithm of $N!$ can be written as:

$$\ln(N!) \approx N\ln(N) - N + \frac{1}{2}\ln(N) + \frac{1}{12N} + \frac{1}{2}\ln(2\pi). \tag{IV.17}$$

We typically deal with over 1000 atoms, and so we would normally keep only the first two terms of this expansion, as the remaining terms make up less than 1% of the total. However, in this particular case, we found that the arithmetic works out to be much simpler if we use the exact factor rather than using this approximation. It is also more accurate, as shortly we will be computing bounds on an expression derived from this in the limit of small atom numbers, where the approximation is not accurate. These bounds would be particularly suspicious as we will allow for the case of 0 or 1 atoms of a particular type. Note that keeping only the first two terms of Equation (IV.17) results in $\ln(0!) \approx 0$, which is actually exact, but $\ln(1!) \approx -1$, which is not only incorrect, but a *decrease* from 0!.

Including the term for the indistinguishability of atoms, we can write out the entropy of atoms of a certain type in our one-way barrier (even though we made some approximations and assumptions about the states of the atoms, we justified that those approximations could be thought of as exact, and so choose to use an equal sign):

$$S_{\mathrm{a}} = NS_0 + N\ln(V) - \ln(N!).$$

The only thing that remains is that we have *two* distinct types of atoms. The atoms that are in the trapped state are distinguishable (optically, in our case, as they absorb a slightly different frequency of light and interact differently with the barrier beam) from the atoms that are not in the trapped state. We make a few more approximations here, and assume that atoms are only pumped to the trapped

atoms are indistinguishable, either classically or quantum mechanically. We will also show that when we do know exactly which atoms are in which state, the entropy change is larger.

state within the trapping volume, and that once trapped, they remain within the trapping volume, with equal probability throughout that volume. We will see later that the entropy change is minimized by the largest possible ratio of untrapped atoms to total number of atoms that see the pumping beam. As the pumping beam decreases that ratio, this assumption is aiming us towards a lower bound for the entropy change, which is our goal. So, we achieve our final entropy value by adding together the individual entropies of the untrapped atoms in the entire volume, and the trapped atoms in the trapping region. This amounts to two separate versions of the above equation. For the untrapped atoms, we replace N with $(1-f)\,N$, and for the trapped atoms, we replace N with fN and V with rV:

$$\begin{aligned} S_{\mathrm{a}} = {} & (1-f)\,NS_0 + (1-f)\,N\mathrm{ln}(V) - \mathrm{ln}([(1-f)\,N]!) \\ & + fNS_0 + fN\mathrm{ln}(rV) - \mathrm{ln}([fN]!) \\ = {} & NS_0 + N\mathrm{ln}(V) - \mathrm{ln}([(1-f)\,N]!) - \mathrm{ln}([fN]!) + fN\mathrm{ln}(r). \end{aligned} \tag{IV.18}$$

We will only be interested in the change in entropy between two states with different numbers of trapped atoms (different values of f). As such, when we take the difference of those two entropy values, the terms that are independent of f will cancel, including the S_0 term representing all the factors that affect the entropy that we do not need to worry about. There are three terms that depend on f. The final term represents a *decrease* in entropy due to atoms being trapped in some fraction r of the total volume. The first two terms represent an *increase* in entropy due to having multiple types of atoms.[31] This increase happens because there is more entropy in a mixture of two types of particles than in the same number of particles of the same type (this, with the second law of thermodynamics, is why mixtures tend to mix).[32]

[31] These terms are actually negative, but larger than the $-\mathrm{ln}(N!)$ that they reduce to in the case where there is only one type of atom ($f = 0$ or $f = 1$). To see this, try subtracting these terms for $f = 0$ from the arbitrary f case. The result is the logarithm of the combinatoric factor $N!/\,(fN)!\,(N-fN)!$. This combinatoric factor can be shown to be larger than 1 whenever fN is an integer larger than 0 and smaller than N by comparing the factors of the numerator with the factors in the denominator. Since this factor is larger than 1, the logarithm is positive, and therefore the entropy for $0 < f < 1$ is larger than the entropy for $f = 0$ or $f = 1$.

[32] We note that without these mixing terms, the entropy can actually be found to decrease. This

At this point, we can make a very quick approximation to show how the entropy increases overall. The assumptions in this approximation are rather questionable, but serve as a quick demonstration how the entropy increase in the pumping beam can overcome the entropy decrease of the atoms. We will continue with our more accurate demonstration after this quick version. The entropy change from having no atoms trapped ($f = 0$) to all atoms trapped ($f = 1$), for the atoms alone, works out to be $N\ln(r)$. We can compute that either from subtracting Equation (IV.18) with $f = 0$ from the same equation with $f = 1$, or by remembering that the entropy, in cases such as these, is related to the logarithm of the available volume (and, since in the initial and final cases, the atoms are all in one state, there is no entropy from having atoms in different states). This entropy change is obviously negative. We can provide a very rough estimate of the pumping beam entropy as follows. For any given atom, the probability of it being in the trapping volume is r. We might thus naïvely expect the probability of trapping that atom to be r, and so assume that it takes $1/r$ pumping beam photons to trap that atom. With N atoms, we might expect it to take roughly N/r atoms to trap all N atoms.[33] Therefore, the final pumping beam might consist of N/r photons, of which N were in a state that means they trapped an atom. The photons are ordered, representing the times at which atoms became trapped, and the number of possible orderings is given by the combinatoric choose function $\binom{N/r}{N} = (N/r)!/N!\,(N/r - N)!$. The entropy of the pumping beam is the logarithm of this, which we approximate using the first two terms of the Stirling approximation in Equation (IV.17) for each of the factorials. The non-logarithm terms cancel, and we are left with:

$$\text{pumping beam entropy} = \frac{N}{r}\ln\left(\frac{N}{r}\right) - N\ln(N) - \left(\frac{N}{r} - N\right)\ln\left(\frac{N}{r} - N\right).$$

is because we can take a limit where a single photon is nearly guaranteed to trap an atom (where almost all the atoms are not trapped). In this limit, there is practically no entropy increase in the pumping beam, while the atomic entropy still decreases because an atom was trapped. With the mixing terms, the atomic entropy increase counters this. With distinguishable atoms, a term very similar to these mixing terms comes in later that prevents a decrease in entropy. This also provides and argument that distinguishable particles would tend to mix as well.

[33] This blatantly ignores the fact that, initially, it is much easier to trap atoms as there are many untrapped atoms in the trapping volume, but the order of magnitude is about right.

Factoring N out of the arguments of the logarithm, we find that the $\ln(N)$ terms also cancel, and the rest reduces to:

$$\text{pumping beam entropy} = -\frac{N}{r}\ln(1-r) + N\ln\left(\frac{1-r}{r}\right).$$

Once we add in the atomic entropy $N\ln(r)$, this combines with the second term to make it a $\ln(1-r)$ term, which combines with the first term to become:

$$\text{total entropy} = -\frac{N}{r}\,(1-r)\ln(1-r).$$

This is N/r multiplied by a factor of the form $-x\ln(x)$, where $x = 1-r$ is between 0 and 1, inclusive. Both of these factors are non-negative, and so in the *total* entropy, the record of when the atoms became trapped in the pumping beam is enough of an entropy increase to counter the decrease in the entropy of the atoms.

We now consider the interaction of the atoms with the pumping beam in a more rigorous fashion. First, we assume the light is composed of discrete particles. Quantum mechanically, we are referring to photons of light, but we can keep this purely classical by referring to very small segments of the light beam, with just enough energy to change the state of an atom (or a small enough amount of light that the probability of changing the state of two atoms is negligible). We will use the term photon for these particles or small chunks of light, with the reservation that we only use it to refer to the smallest amount of light required to shift the state of an atom, and not as a full quantum-mechanical quantization of the electromagnetic field.

Just as the atoms have two possible states, the photon must also have two possible states. In our case, the two states are different frequencies. In general, the initial photon is capable of being absorbed by untrapped atoms, but not by trapped atoms. If we simply state that the process is irreversible, then we are leaving out an important entropy increase, and our result will not necessarily show a net increase in entropy. We must be careful to include any possible increase in entropy by reducing everything to reversible processes. When a photon changes an atom from the untrapped state to the trapped state, the photon is actually absorbed and re-emitted. Assuming all processes are reversible, that means the reversed process can also occur: A re-emitted photon is capable of untrapping a trapped atom. Since we need a situation where the pumping photons cannot untrap atoms, we must have the re-emitted photon be somehow different from the absorbed photon.[34] The difference could be in polarization, frequency, direction, or some other property (like mass, as nothing in this description requires our "photons" to be actual light particles). Our input beam of light consists entirely of identical photons in the state that can trap atoms. For each atom trapped, the output beam will have one of those photons replaced with a photon in the other state. The position of these replaced photons gives a record of when atoms were trapped, and carries some information about the initial distribution of atoms in the initial volume. The mixing of these two states of photons increases the entropy of the pumping beam and, as we will show, when added to the net change of the atomic entropy, results in a net *increase* (or no change, in certain limits) in the total entropy. In other words, the reduction of the entropy of the atoms (by trapping them in a smaller volume) is countered by an *increase* in entropy of the pumping beam used to trap the atoms, as the pumping beam carries away the extra position information of the atoms. This is essentially dissipation, although here we are not necessarily carrying energy away, but we are carrying information away.

Let us watch a single photon as it passes through the trapping volume, or some subset of that region which we will call the interaction volume. Our trap is narrow, so we will assume the photon crosses the trap quickly enough that the number of atoms within the interaction volume do not change during the passing. Fur-

[34] We note that there are reactions that utilize a catalyst to occur in only one direction, such as manganese dioxide enhancing the decomposition of hydrogen peroxide, or the CNO cycle in stellar fusion. These are examples where a "photon" (the catalyst) causes an irreversible change in the "particle" (the hydrogen peroxide, or the hydrogen that gets fused in our examples), but the "photon" *is* the same as before the change. To be irreversible, they must have an entropy increase associated with them. In these cases, the reactions release energy as a random kick to the reactants. The reactions are technically reversible in that if, right after the reaction, we reversed the directions of everything, they would recombine. However, that energy kick is quickly lost to the environment as the particles collide, and after that, the particles no longer have the energy necessary to recombine in the same way. That loss of energy to the environment (dissipation) is the entropy increase that makes the reaction irreversible. In our atom–photon interactions, we have already explicitly stated that we do not lose energy to the environment (and no random kicks), and so we require some other mechanism to result in irreversibility.

thermore, we will assume that the photon interacts with all the atoms within the trapping volume at the same time as it passes through (as light travels at a very high velocity, it tends to be hard to localize it in space). Really, our argument is that the probability of interaction with an atom is a constant, and we will use that to put a bound on the probability of an interaction. As we do not know exactly which atoms the photon will encounter and in which order, we argue that the final probability of an interaction occurring, once we account for any possible reversals (after trapping an atom, the re-emitted photon can untrap an atom), is proportional to the *ratio* of the number of trappable atom to the total number of atoms in the interaction volume (plus one, as will be described shortly). We suspect that that ratio is the upper limit for the probability of trapping an atom. We then average over the expected number of atoms within the interaction volume and compute an upper bound for the trapping probability. Our assumptions are one way to compute an upper bound for the trapping probability, but we would argue that other situations would present the same upper bound.

We can either pretend the photon is passing through the trap, or simply spending time in the trap region, and compute the probability of trapping an atom from that. We choose a spatial representation, and assign the coordinate z for how far through the trap the photon has travelled (or what fraction of the length of the photon has passed through the trap, since the trap is narrow). As we have argued, the probability of an interaction is proportional to the number of atoms in the correct state being within the interaction volume. The proportionality constant includes such things as beam intensity, the size of the interaction volume, and other things, but, most importantly, must be the same for trapping as untrapping an atom (after all, the process should be reversible).

Using $\mathcal{P}(z)$ to denote the probability of the photon being in the state that untraps atoms (meaning it has already trapped an atom), the probability of it being in the other state is $1 - \mathcal{P}(z)$. The change in that probability as the photon passes through some thin slice $\mathrm{d}z$ is the probability that the photon traps an atom (times the probability that it was in the correct state for that, $1 - \mathcal{P}(z)$), putting it in the state that untraps atoms, minus the probability that it is already in that state

($\mathcal{P}(z)$) and it untraps an atom, putting in the state that traps atoms. All of this is also proportional to the thickness of this slice, dz. The only non-common factors between those two processes are the number of untrapped atoms in the interaction volume, which we call u, and the number of trapped atoms in the interaction volume, which we call t. Note that if the photon has trapped an atom, and we are computing the part where the photon might untrap an atom, the number of trapped atoms is the number of previously trapped atoms plus the extra atom that was trapped:

$$\mathrm{d}\mathcal{P}(z) \propto (u\,\mathrm{d}z)\,(1-\mathcal{P}(z)) - ((t+1)\ \mathrm{d}z)\,\mathcal{P}(z).$$

We can absorb the proportionality constant into the scaling of z, and treat this as an equation, rather than just a proportionality relation. This changes z into a length-scale for optical interactions (the optical depth), which we can rewrite into differential-equation form:

$$\frac{\mathrm{d}\mathcal{P}(z)}{\mathrm{d}z} = u\,(1-\mathcal{P}(z)) - (t+1)\,\mathcal{P}(z) = u - (u+t+1)\,\mathcal{P}(z). \tag{IV.19}$$

Keeping our assumption that u and t do not change while the photon passes through, this equation is easy to solve:

$$\begin{aligned}
\frac{\mathrm{d}\mathcal{P}(z)}{\mathrm{d}z} + (u+t+1)\,\mathcal{P}(z) &= u \\
\frac{\mathrm{d}\mathcal{P}(z)}{\mathrm{d}z}e^{(u+t+1)z} + (u+t+1)\,\mathcal{P}(z)e^{(u+t+1)z} &= ue^{(u+t+1)z} \\
\frac{\mathrm{d}}{\mathrm{d}z}\left[\mathcal{P}(z)e^{(u+t+1)z}\right] &= ue^{(u+t+1)z} \\
\mathcal{P}(z)e^{(u+t+1)z} &= \frac{u}{u+t+1}e^{(u+t+1)z} + C \\
\mathcal{P}(z) &= \frac{u}{u+t+1} + Ce^{-(u+t+1)z}.
\end{aligned}$$

Our photon is initially in the state that traps atoms, so $\mathcal{P}(z=0)=0$. We can solve for the constant of integration and get a solution for the probability that the photon has trapped an atom (net) at any point z:

$$\mathcal{P}(z) = \frac{u}{u+t+1}\left(1-e^{-(u+t+1)z}\right). \tag{IV.20}$$

Note that $\mathcal{P}(z)$ starts at 0 and increases monotonically towards a maximum limit of $u/\,(u+t+1)$ as the optical density increases. Our final entropy value will decrease as $\mathcal{P}(z)$ increases, indicating that the entropy increase is minimized for a large optical depth, and so we will use this maximum value for our lower bound on entropy change. In any case where the probability of a transition is arbitrarily weak, we can reduce this probability to zero. However, even if we change the details of how this transition occurs, if the probability is large, a transition will occur rapidly. Since the process must be inherently reversible, the reverse will also occur rapidly. The transition will happen back and forth, and since we do not know where it stops, the probability of a transition having occurred at that point is the relative speed of the trapping transition to the speed of a full trap/untrap cycle (proportional to the number of atoms, but counting one twice, as once an atom is trapped, that one also counts as a trapped atom to increase the likelihood of untrapping an atom). The only factor that changes the speed of the transitions is the number of atoms available, which matches the value we have calculated for this specific set of assumptions.

There is a subtlety here that we will elaborate upon. We have two parts to our system (atoms and the photon). After the interaction, each system is in one of two collections of states (whether or not another atom was trapped). However, the two states are related, in that if an atom was trapped, the photon is in the corresponding state, and vice versa. This means that the system entropy is increased by us not knowing what happened, but that entropy increase is not specifically attached to either the atoms or the photon. If the atoms and the photon were each *independently* in one of two states (or collections of states), then we would just add entropies associated with each one. However, once the collection of the atoms is determined, the state of the photon is *also* determined, so instead of adding an entropy term for the atoms as well as the photon, there is just a single addition. However, the probabilities work out such that, as the process continues, the atoms eventually end up in a known state (all trapped) with very high probability, for an apparent entropy decrease. The stream of photons, however, keeps that entropy increase associated with not knowing when that happens. That entropy increase is related to when exactly the individual atoms passed through the trapping region, and so, in a sense, is a partial record of where the atoms were in the trap initially. If the atom locations had been known in advance, the initial atom volume for each atom we would have used for the entropy would need to be smaller than the trapping region, and so the initial state would have the same (or less) volume than the final state, and there is no entropy decrease. If we do *not* know the locations well enough in advance, then even though we do after a sufficient period of time, the same amount (or more) information about where the atoms were initially is transferred into the photon stream, which more than compensates for the decrease in atomic entropy.

We now show that the entropy change decreases as $\mathcal{P}$, the value of $\mathcal{P}(z)$ after the photon exits the trapping region, increases. This essentially is a statement that wasting photons so that it takes more of them, on average, to trap an atom, increases the entropy. Before the interaction, the photon is in a known state (the state that can trap an atom), and the atoms have the entropy given by Equation (IV.18). After the interaction, with probability $\mathcal{P}$, the number of trapped atoms has increased by one (f has increased by $1/N$), and with probability $1-\mathcal{P}$, the atomic entropy is unchanged.

To simplify, say the initial entropy before the interaction (given by Equation (IV.18)) is S_0. With probability $1-\mathcal{P}$, the entropy is the same, but with probability $\mathcal{P}$, the entropy is that given by Equation (IV.18) with a substitution $f \to f+1/N$, which we will call S_{trap}. The states associated with these two entropies are distinct, with no overlap (as the number of trapped atoms is different for the two sets). We can use this to compute an entropy for this case where we do not even know which set of states the system is in, using the same method we used to derive Equation (IV.13). The catch is that not all the states are equally likely (some have an extra weighting factor of $\mathcal{P}$, and some have $1-\mathcal{P}$). The entropy is then computed from the definition in Equation (IV.11):

$$S = -\sum_{i=1}^{N} \mathcal{P}_i \text{ln}(\mathcal{P}_i).$$

The sum is over the states for *both* S_0 and S_{trap}. When i refers to a state corresponding to S_0, $\mathcal{P}_i$ is the probability used in computing S_0 multiplied by $1-\mathcal{P}$ (the probability of being in that set of states). For the remainder of the states (those corresponding to $S_{\rm trap}$), the probabilities are the same as those used in $S_{\rm trap}$, multiplied by the probability of being in that set of states ($\mathcal{P}$). Therefore, we can break the sum up into those two sets, and explicitly write the probabilities in terms of the ones used to initially compute the entropies by adding the $\mathcal{P}$ and $1-\mathcal{P}$ factors (using the notation $i \in \{S_0\}$ to mean i sums over the states corresponding to S_0):

$$\begin{aligned}
S &= -\sum_{i\in\{S_0\}} (1-\mathcal{P})\,\mathcal{P}_i \ln((1-\mathcal{P})\,\mathcal{P}_i) - \sum_{i\in\{S_{\rm trap}\}} \mathcal{P}\mathcal{P}_i \ln(\mathcal{P}\mathcal{P}_i) \\
&= -(1-\mathcal{P}) \sum_{i\in\{S_0\}} \mathcal{P}_i \left\{\ln(1-\mathcal{P}) + \ln(\mathcal{P}_i)\right\} - \mathcal{P} \sum_{i\in\{S_{\rm trap}\}} \mathcal{P}_i \left\{\ln(\mathcal{P}) + \ln(\mathcal{P}_i)\right\} \\
&= -(1-\mathcal{P}) \sum_{i\in\{S_0\}} \mathcal{P}_i \ln(1-\mathcal{P}) - (1-\mathcal{P}) \sum_{i\in\{S_0\}} \mathcal{P}_i \ln(\mathcal{P}_i) \\
&\quad -\mathcal{P} \sum_{i\in\{S_{\rm trap}\}} \mathcal{P}_i \ln(\mathcal{P}) - \mathcal{P} \sum_{i\in\{S_{\rm trap}\}} \mathcal{P}_i \ln(\mathcal{P}_i) \\
&= -(1-\mathcal{P})\ln(1-\mathcal{P}) + (1-\mathcal{P})\,S_0 - \mathcal{P}\ln(\mathcal{P}) + \mathcal{P}S_{\rm trap}.
\end{aligned}$$

In the last equality, we used the facts that the sum of $\mathcal{P}_i$ over a given set of states must be 1, and the sum of $-\mathcal{P}_i\ln(\mathcal{P}_i)$ over a given set is the entropy for that set. We can therefore write the final entropy as:

$$S = (1-\mathcal{P})\,S_0 + \mathcal{P}S_{\rm trap} - (1-\mathcal{P})\ln(1-\mathcal{P}) - \mathcal{P}\ln(\mathcal{P}). \qquad \text{(IV.21)}$$

This is simply a weighted average of the entropies, plus two extra entropy terms (that are non-negative) related to the uncertainty in whether an atom was trapped or not. While this excess entropy is shared between the atoms and the pumping beam photon, we can think of this term as an increase in the entropy of the pumping beam. While there is initially uncertainty in whether atoms are trapped or not, the atoms eventually end up in the system where all the atoms are trapped. However, the entropy is essentially "carried away" by the pumping beam.

Now, we need only compute the change in entropy, and show that it is never negative. We do this by showing that the change in entropy for every photon increases the entropy. Even though we might conceivably compute the entropy relative to a different zero with each photon (since after each atom is trapped, we might need to change our "volume" distribution), the fact that entropy does not decrease with each step means that the sum of all the entropy changes over many photons also must not decrease. We start with the assumption that the atoms are evenly spread out within the volume (or, at least, the probability distribution for the atoms is evenly spread). We discussed earlier how the volume could be remapped to satisfy this requirement. Furthermore, we would actually expect the entropy at a beginning step to have an uncertain number of trapped atoms (up until we measure the number that have been trapped). In a manner similar to the derivation of Equation (IV.21), we would expect the entropy afterwards to be a weighted sum of the entropies, plus some extra entropy for the extra states that entails. By assuming we are starting with a known number of trapped atoms, we are actually underestimating the entropy change. If we find that the net entropy change for a well-known number of trapped atoms initially is non-negative, then the entropy change for a spread of trapped-atom numbers should be larger, and therefore also non-negative. Every time an atom is trapped, the distribution of atoms changes some. We could simply wait long enough for the distribution to become even again. We could also re-distribute our volume mapping, which changes the entropy for the next step, but since we are only concerned with showing that each incremental change is non-negative, this is not a problem. For this incremental step, we are using S_0 to represent the entropy before a photon passes through the atoms, and S_{trap} to represent the entropy afterwards *if* an extra atom gets trapped by the photon.

The net entropy after the interaction is given by Equation (IV.21), so the change in entropy after this one photon passes through is simply S_0 subtracted from that:

$$\begin{aligned}\Delta S &= (1-\mathcal{P})\, S_0 + \mathcal{P} S_{\text{trap}} - (1-\mathcal{P})\ln((1-\mathcal{P})) - \mathcal{P}\ln(\mathcal{P}) - S_0 \\ &= \mathcal{P}\,(S_{\text{trap}} - S_0) - (1-\mathcal{P})\ln(1-\mathcal{P}) - \mathcal{P}\ln(\mathcal{P}).\end{aligned}$$

Now, we substitute for S_0 (given by Equation (IV.18)) and S_{trap} (same equation with f increased by $1/N$). As we mentioned earlier, we see that the change in entropy depends only on the difference between two versions of Equation (IV.18) that differ only in the value of f, and so all the terms independent of f cancel.[35]

$$S_0 = NS_0 + N\ln(V) - \ln([(1-f)\,N]!) - \ln([fN]!) + fN\ln(r) \tag{IV.22}$$

$$\begin{aligned} S_{\text{trap}} = {} & NS_0 + N\ln(V) \qquad\qquad\qquad\qquad . \\ & -\ln\left(\left[\left(1-f-\frac{1}{N}\right)N\right]!\right) - \ln\left(\left[\left(f+\frac{1}{N}\right)N\right]!\right) \\ & +\left(f+\frac{1}{N}\right)N\ln(r) \end{aligned} \tag{IV.23}$$

Subtracting these and combining terms yields:

$$\begin{aligned} S_{\text{trap}} - S_0 = {} & NS_0 + N\ln(V) \\ & -\ln([(N-fN-1)]!) - \ln([(fN+1)]!) \\ & +(fN+1)\ln(r) \\ & -NS_0 - N\ln(V) + \ln([(N-fN)]!) + \ln([fN]!) - fN\ln(r) \end{aligned} \tag{IV.24}$$

$$\begin{aligned} = {} & -\ln([(N-fN-1)]!) + \ln([(N-fN)]!) \\ & -\ln([(fN+1)]!) + \ln([fN]!) \\ & +\ln(r). \end{aligned} \tag{IV.25}$$

The first two lines are of the form $\ln((n+1)!) - \ln(n!)$, which is equal to the logarithm of $(n+1)!/n! = (n+1)$:

$$\begin{aligned} S_{\text{trap}} - S_0 = {} & \ln(N-fN) \\ & -\ln(fN+1) \\ & +\ln(r). \end{aligned} \tag{IV.26}$$

[35] If the atoms were distinguishable, then the factorial terms, which came from the indistinguishability of atoms, would be missing. In that case, the difference of the two entropies would simply be $fN\ln(r)$. However, if we know exactly which atoms are in which state initially, we still do not know which of the untrapped atoms got trapped (if one did get trapped), so there are actually $N-fN$ (the number of untrapped atoms) times as many states corresponding to S_{trap}. Therefore, we would need to add a $\ln(N-fN)$ term to S_{trap}. Thus, we would get a very similar difference to the indistinguishable case, but missing the $-\ln(fN+1)$, which means this difference is *larger* in the distinguishable atom case. As described previously, if we do not know which atoms are in which state initially, then the resulting increases to the before and after entropies work out to give the exact same result as the indistinguishable atom method.

Substitute this difference back into our incremental entropy change for the entire system:

$$\Delta S = \mathcal{P}\left\{\ln(N - fN) - \ln(fN + 1) + \ln(r)\right\} \qquad \text{(IV.27)}$$
$$- (1 - \mathcal{P})\ln(1 - \mathcal{P}) - \mathcal{P}\ln(\mathcal{P}).$$

As a quick note, this equation is only valid if there is at least one atom left to trap, since we were combining terms where we either did or did not trap an atom. Thus, fN, the number of trapped atoms *before* the interaction, is between zero and $N - 1$, inclusive. The quantity $N - fN$, which represents the number of untrapped atoms before the interaction, must be between *one* and N, inclusive. From Equation (IV.20), we see that if there are no untrapped atoms ($u = 0$), then $\mathcal{P}$ is zero. Substituting zero for $\mathcal{P}$ sets the entire entropy equation to zero, as can be seen by direct substitution, with the exception of the last term. The last term continuously approaches zero as $\mathcal{P}$ goes to zero.[36] This makes sense because the original definition of entropy is a sum of terms of that form, and the addition of inaccessible states (states with probability 0) should not change the entropy.

Now we are left trying to show that Equation (IV.27) is non-negative. We will treat $\mathcal{P}$ as a positive quantity, since we know the entropy does not change (and therefore does not decrease) if $\mathcal{P}$ is zero. Now that we have ruled out the $\mathcal{P} = 0$ case, we can factor $\mathcal{P}$ out without changing the sign of the quantity. We are now trying to show that this new quantity is non-negative:

$$\frac{\Delta S}{\mathcal{P}} = \left\{\ln(N - fN) - \ln(fN + 1) + \ln(r)\right\} \qquad \text{(IV.28)}$$
$$- \left(\frac{1}{\mathcal{P}} - 1\right)\ln(1 - \mathcal{P}) - \ln(\mathcal{P}).$$

We will be finding the minimum value of this quantity by showing that this function decreases as $\mathcal{P}$ increases, and then we will show that this quantity is non-negative even with the maximum possible $\mathcal{P}$.

[36] This can be seen using l'Hôpital's Rule on the quantity $\ln(\mathcal{P})/\left(1/\mathcal{P}\right)$, where both the numerator and denominator approach infinity as $\mathcal{P} \to 0$. Taking the derivative of the numerator and denominator simplifies to $-\mathcal{P}$, which tends to 0 as $\mathcal{P} \to 0$.

The first derivative of Equation (IV.28) with respect to $\mathcal{P}$ is:

$$\begin{aligned}\text{First derivative} &= \frac{1}{\mathcal{P}^2}\ln(1-\mathcal{P}) + \left(\frac{1}{\mathcal{P}} - 1\right)\frac{1}{1-\mathcal{P}} - \frac{1}{\mathcal{P}} \\ &= \frac{1}{\mathcal{P}^2}\left\{\ln(1-\mathcal{P}) + \mathcal{P}\left(1-\mathcal{P}\right)\frac{1}{1-\mathcal{P}} - \mathcal{P}\right\} \\ &= \frac{\ln(1-\mathcal{P})}{\mathcal{P}^2}. \end{aligned} \tag{IV.29}$$

Since $\mathcal{P}$ is a probability, it is non-negative with an upper bound of one. This means the first derivative of Equation (IV.28) is the logarithm of a non-negative number of at most one, which is either negative or zero. Since we are trying to get the minimum possible value, we therefore need to pick the largest allowed value for $\mathcal{P}$. This brings us back to our equation for $\mathcal{P}$, Equation (IV.20) (where z is set to the optical length of the pumping beam path through the trap), which has a maximum value of $u/\left(u+t+1\right)$. At the start of the photon interaction we are working with, we stated that we were assigning volume such that the atoms were equally spread throughout the volume. This is the case where the entropy associated with trapping an extra atom has the largest decrease, since if the atoms are more localized, then the effective volume that they are in is smaller. If the initial "volume" is smaller, than the decrease in volume associated with trapping an atom is less significant.

We have an upper limit for the maximum value of $\mathcal{P}$ given that we know the numbers atoms in each state in the trapping region of the beam. However, since our interaction time is short, we take a snapshot of the current numbers of atoms within the interaction volume, which can have fluctuations. We will then show that $\mathcal{P}$ is maximized by the largest possible interaction volume, which makes sense because by interacting with the most atoms, we minimize the chances of wasting photons. This is easy to show in the long-interaction case, where we just simply insert average values in for u and t in our limit for $\mathcal{P}$, and take a derivative with respect to the interaction volume. Furthermore, we will show that even with fluctuations, once we are using the largest possible interaction volume, the entropy is bounded by the case where we use the average number within the volume. That means our results are also valid in the limit of long interaction time, where the photon effectively interacts with the atoms many times, effectively scaling the number of atoms up by some large factor (but keeping the proportionalities). That, in turn, is why the case with averages works, as having much larger numbers makes the fluctuations about the averages negligible.

We know an upper bound for the interaction probability given a fixed number of each type of atom within the interaction volume, u and t. If we knew the actual probability, then the final probability would be that averaged over all realizations of u and t. However, that average is just a weighted sum of probabilities (where all terms are non-negative), and so if we take the same weighted sum of the upper bounds, the result is an upper bound for the interaction probability taking into account all fluctuations in u and t.

We call the interaction volume (divided by the total volume) $\mathfrak{t}$, the total number of untrapped atoms U, and the total number of trapped atoms T. Since the probability of each atom being within the interaction volume is independent of the number of atoms already in there, we can treat u and t as binomially distributed random variables. The probability of a particular untrapped atom being in the interaction volume is simply the ratio of the interaction volume to the total volume, $\mathfrak{t}$. The probability of a particular trapped atom is that ratio of the interaction volume to the trapped volume (the region that the trapped atom is confined in), which is $\mathfrak{t}/r$. As discussed, our upper bound for trapping an atom, $\mathcal{P}$, is then the average of $u/\left(u+t+1\right)$ over the two binomial distributions:

$$\mathcal{P} = \sum_{u=0}^{U}\sum_{t=0}^{T}\binom{U}{u}\binom{T}{t}\mathfrak{t}^{u}\left(1-\mathfrak{t}\right)^{U-u}\left(\frac{\mathfrak{t}}{r}\right)^{t}\left(1-\frac{\mathfrak{t}}{r}\right)^{T-t}\frac{u}{u+t+1}. \tag{IV.30}$$

Rather than compute this expectation value, we will be using various bounds on it.

Intuitively, we might expect that the interaction probability in Equation (IV.30) is largest with the largest possible interaction volume ($\mathfrak{t} \to r$, since so do not want to be "trapping" atoms outside the trap volume), where we interact with every atom within the trapping region. Certainly, this maximizes the number of untrapped atoms we have to interact with, which raises our bound on $\mathcal{P}$, but it also increases the number of trapped atoms, which decreases our bound on $\mathcal{P}$. Since we interact first with the untrapped atoms, we might expect maximizing u to be more important than minimizing t, and this does turn out to be the case. This greatly simplifies finding bounds on $\mathcal{P}$, because we can simply take the $\mathfrak{t} \to r$ limit, in which case only the $t = T$ term of the t sum is non-zero. In this limit, where the interaction volume *is* the trapping volume, *all* the trapped atoms are within the interaction volume, with unity probability, and so we can simply replace t with T, and only worry about the u average.

Our proof that $\mathcal{P}$ is maximized by the largest trapping volume, however, is unwieldy. Originally, we proved this for the slightly larger bound $u/\left(u + t\right)$, before we realized that we could count one particular atom twice.[37] That version reduced to a surprisingly simple form, which leads us to suspect that there is a cleaner proof, but time pressures prevented us from playing with this further. We begin by taking the derivative of Equation (IV.30) with respect to the interaction volume. To simplify, we use the substitution $\mathfrak{d} = 1 - \mathfrak{t}/r$ for every occurrence of $1 - \mathfrak{t}/r$ and $1/r$ (which comes from the derivative of $\mathfrak{d}$ with respect to $\mathfrak{t}$). Our result is this:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}}\mathcal{P} = & \sum_{u=0}^{U}\sum_{t=0}^{T}\binom{U}{u}\binom{T}{t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1}\mathfrak{d}^{T-t-1}\left(1-\mathfrak{d}\right)^{t} \\ & \times\left[\left(u - \mathfrak{t}U + T - \mathfrak{t}T\right)\mathfrak{d} - \left(1-\mathfrak{t}\right)\left(T-t\right)\right]\frac{u}{u+t+1}.\end{aligned} \tag{IV.31}$$

We note that we took a derivative of what is essentially a polynomial in $\mathfrak{t}$, and so the apparent singularities at $\mathfrak{t} = 1$ (and $u = U$) or $\mathfrak{d} = 0$ (and $t = T$) actually disappear when the appropriate u or t substitution is made throughout, and the result simplified.

Our next step is to re-write Equation (IV.31) as a polynomial in $\mathfrak{d}$. Essentially, we expand the power of $(1 - \mathfrak{d})$ using a binomial expansion, and then swap that sum with the t sum. This involves a little trickery with the sum limits, and we find it helps to sketch the region of values that the sums cover on a graph. The region

[37] The proof is almost identical to this one, with simply a replacement of the denominator. However, since $u/\left(u + t\right)$ has a singularity at $u = t = 0$, the proof required several special cases. We handled these initially by excluding the $u = 0$ cases from the sum. This is justified because the terms are all 0, trivially when t is non-zero, and because if $u = t = 0$, then there are no atoms at all, and so the probability of a transition is 0. We will briefly footnote where that proof differed from this one (once the sum limits and the denominator have been switched), but, since it was more cumbersome, we will skip most of the details.

is a right-triangle, and so we can pick either axis as our independent axis, and pick the limits of the other variable as a function of that. Having the extra $\mathfrak{d}$ in the brackets requires splitting this into two sums, as the p limits end up with different limits as a result:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}}\mathcal{P} = \quad & \sum_{p=0}^{T-1}\sum_{u=1}^{U}\sum_{t=T-p-1}^{T-1}\binom{U}{u}\binom{T}{t}\binom{t}{p-T+t+1}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \qquad \times\left(-1\right)^{p-T+t}\left[\left(1-\mathfrak{t}\right)\left(T-t\right)\right]\frac{u}{u+t+1}\mathfrak{d}^{p} \\ & +\sum_{p=0}^{T}\sum_{u=1}^{U}\sum_{t=T-p}^{T}\binom{U}{u}\binom{T}{t}\binom{t}{p-T+t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \qquad \times\left(-1\right)^{p-T+t}\left[\left(u-\mathfrak{t}U\right)+\left(1-\mathfrak{t}\right)T\right]\frac{u}{u+t+1}\mathfrak{d}^{p}.\end{aligned} \tag{IV.32}$$

While these two sums can be combined directly, we found the result is simpler if we first employ a trick. In the first sum, if we shift the t summation variable to $t+1$ (replacing $t+1$ with t), then the t sum has the same limits in each of the sums. We can fix the denominator of the $u/\left(u+t+1\right)$ by simultaneously shifting the u summation variable to $u-1$. Taking advantage of the fact that the terms where $u=0$ are all zero because of the u in the numerator of $u/\left(u+t+1\right)$, we can drop those terms to avoid having an $u=-1$ in the sum limits after shifting u. After performing these shifts, we can then rearrange some factors in each sum, using these identities which are easily verified by writing out the combinatoric factors as fractions of factorials:

$$\begin{aligned}\binom{U}{u+1}\left(u+1\right) &= \binom{U}{u}\left(U-u\right) \\ \binom{T}{t-1}\binom{t-1}{p-T+t}\left(T-t+1\right) &= \binom{T}{p}\binom{p}{T-t}\left(T-p\right) \\ \binom{T}{t}\binom{t}{p-T+t} &= \binom{T}{p}\binom{p}{T-t}.\end{aligned}$$

The result of these manipulations is:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}}\mathcal{P} = \quad & \sum_{p=0}^{T-1}\sum_{u=0}^{U-1}\sum_{t=T-p}^{T}\binom{U}{u}\binom{T}{p}\binom{p}{T-t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \qquad \times\left(-1\right)^{p-T+t}\left[-\mathfrak{t}\left(U-u\right)\left(T-p\right)\right]\frac{1}{u+t+1}\mathfrak{d}^{p} \\ & +\sum_{p=0}^{T}\sum_{u=0}^{U}\sum_{t=T-p}^{T}\binom{U}{u}\binom{T}{p}\binom{p}{T-t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \qquad \times\left(-1\right)^{p-T+t}\left[u\left(u-\mathfrak{t}U\right)+u\left(1-\mathfrak{t}\right)T\right]\frac{1}{u+t+1}\mathfrak{d}^{p}.\end{aligned}\tag{IV.33}$$

In this form, we can extend the p upper limit in the first sum to T (as the $(T-p)$ factor makes that term zero), and we can extend all the u limits to be from 0 to U (the extra terms are all zero due to certain factors within the brackets).[38] After some algebraic manipulation, we arrive with this result,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}}\mathcal{P} = & \sum_{p=0}^{T}\sum_{u=0}^{U}\sum_{t=T-p}^{T}\binom{U}{u}\binom{T}{p}\binom{p}{T-t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \times\left(-1\right)^{p-T+t}\left[\left(u+T\right)\left(u-\mathfrak{t}U\right)+\mathfrak{t}p\left(U-u\right)\right]\frac{1}{u+t+1}\mathfrak{d}^{p},\end{aligned}\tag{IV.34}$$

which we will use in an induction argument.

For brevity and clarity, we define a shorthand notation for the coefficient of the $\mathfrak{d}^{p}$ term in Equation (IV.34):[39]

$$\begin{aligned}\mathbf{C}(T,p) := & \sum_{u=0}^{U}\sum_{t=T-p}^{T}\binom{U}{u}\binom{T}{p}\binom{p}{T-t}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1} \\ & \times\left(-1\right)^{p-T+t}\left[\left(u+T\right)\left(u-\mathfrak{t}U\right)+\mathfrak{t}p\left(U-u\right)\right]\frac{1}{u+t+1}.\end{aligned}\tag{IV.35}$$

Technically, we are not showing the full functional dependence of the coefficient (leaving off $\mathfrak{t}$, for example), but these extra dependencies will not matter for our

[38] In the similar proof using $u/\left(u+t\right)$ instead of $u/\left(u+t+1\right)$, the second sum had a lower limit of $u=1$ instead of $u=0$, but the first sum, due to the shift, had a lower limit of $u=0$. We chose to pull out the $p=T$ terms of the second sum, and deal with those separately. For the terms with $p<T$, the limits on the t sum excluded $t=0$, so we could add the $u=0$ terms on without any divide-by-zero problems arising (the numerator is zero for those terms).

[39] In the similar proof using $u/\left(u+t\right)$ instead of $u/\left(u+t+1\right)$, this was the coefficient for $p<T$ (except for the denominator being $u+t$ in that case), but the lower limit on the u sum was 1, not 0, when $p=T$.

argument. We now plan to use an induction-like recursion relation on these coefficients, where we relate each coefficient to terms with smaller p values. To do so, we start with a well-known relation for binomial coefficients (easily shown by breaking the coefficients up into fractions of factorials): $\binom{p}{T-t} = \binom{p-1}{T-t} + \binom{p-1}{T-t-1}$. Using this expansion, then, we can write the two binomial coefficients involving p in Equation (IV.35) in terms of coefficients with smaller p, where we simply pulled a few factors out from the other binomial coefficient (we will assume that $p > 0$ for this):

$$\binom{T}{p}\binom{p}{T-t} = \frac{T}{p}\binom{T-1}{p-1}\binom{p-1}{T-1-t} + \frac{T-(p-1)}{p}\binom{T}{p-1}\binom{p-1}{T-t}. \tag{IV.36}$$

If we insert Equation (IV.36) into Equation (IV.35), and split certain other occurrences of T and p into $(T-1)+1$ and $(p-1)+1$, respectively, we get the following relation:

$$\begin{aligned}
\mathbf{C}(T,p) = \quad & \sum_{u=0}^{U}\sum_{t=T-p}^{T-1}\binom{U}{u}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1}\frac{T}{p}\binom{T-1}{p-1}\binom{p-1}{T-1-t} \\
& \times(-1)^{(p-1)-(T-1)+t}\left[(u+T-1)\left(u-\mathfrak{t}U\right)+\mathfrak{t}\left(p-1\right)\left(U-u\right)\right]\frac{1}{u+t+1} \\
& +\sum_{u=0}^{U}\sum_{t=T-p}^{T-1}\binom{U}{u}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1}\frac{T}{p}\binom{T-1}{p-1}\binom{p-1}{T-1-t} \\
& \qquad\times(-1)^{(p-1)-(T-1)+t}\left[\left(u-\mathfrak{t}U\right)+\mathfrak{t}\left(U-u\right)\right]\frac{1}{u+t+1} \\
& -\sum_{u=0}^{U}\sum_{t=T-p+1}^{T}\binom{U}{u}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1}\frac{T-(p-1)}{p}\binom{T}{p-1}\binom{p-1}{T-t} \\
& \qquad\times(-1)^{(p-1)-T+t}\left[(u+T)\left(u-\mathfrak{t}U\right)+\mathfrak{t}\left(p-1\right)\left(U-u\right)\right]\frac{1}{u+t+1} \\
& -\sum_{u=0}^{U}\sum_{t=T-p+1}^{T}\binom{U}{u}\mathfrak{t}^{u-1}\left(1-\mathfrak{t}\right)^{U-u-1}\frac{T-(p-1)}{p}\binom{T}{p-1}\binom{p-1}{T-t} \\
& \qquad\times(-1)^{(p-1)-T+t}\left[\mathfrak{t}\left(U-u\right)\right]\frac{1}{u+t+1}.
\end{aligned} \tag{IV.37}$$

We modified the limits of the t sum to avoid the cases where the binomial coefficients from Equation (IV.36) would be zero. The first and third sums in Equation (IV.37) are proportional to $\mathbf{C}(T-1,p-1)$ and $\mathbf{C}(T,p-1)$, respectively. The second and fourth sums actually cancel. To show this, first note that the factor in brackets in the second sum reduces to $u\,(1-\mathfrak{t})$, which is zero when $u=0$, so we drop that term of the u sum (likewise, we can drop the $u=U$ term of the fourth sum). We can include the $(1-\mathfrak{t})$ in the brackets of the second sum with the $(1-\mathfrak{t})^{U-u-1}$ factor, and combine the u with the binomial coefficients using this identity (easily shown by expanding as factorials): $\binom{U}{u}u = \binom{U}{u-1}(U-(u-1))$. After making these changes, we then shift the summation variables of the second sum to $u-1$ and $t+1$, and then this sum becomes term-for-term identical with the fourth sum except they are negatives of each other.

The above argument proves the following recursion relation:[40]

$$\mathbf{C}(T,p) = \frac{T}{p}\mathbf{C}(T-1,p-1) - \frac{T-(p-1)}{p}\mathbf{C}(T,p-1). \tag{IV.38}$$

This recursion relation is where the proof using $u/\,(u+t+1)$ reduces to an impressively simple result. For $p=0$, the t sum has only one term, $t=T$. Making those two substitutions in Equation (IV.35), the numerator is proportional to $u+T$, which cancels with the denominator (which is $u+t$ in the other proof). What is left is essentially a binomial average of $u-\mathfrak{t}U$, divided by $\mathfrak{t}$ and $(1-\mathfrak{t})$. Since, as mentioned before, we essentially took the derivative of a polynomial in $\mathfrak{t}$, these divisions should not actually create singularities. Indeed, if we check the terms where we have either $\mathfrak{t}^{-1}$ or $(1-\mathfrak{t})^{-1}$, we see that the $u-\mathfrak{t}U$ cancels those factors in those terms. Since the average of u is $\mathfrak{t}U$, this binomial average is identically zero. Here is where it is important to have kept track of the problem where $u=t=0$, as we canceled the singularity where $p=T=0$. Handled properly, the u sum in Equation (IV.35) should have a lower limit of 1, not 0, when $p=T$. By the same argument

[40] This recursion relation also holds for the proof with $u/\,(u+t)$ instead of $u/\,(u+t+1)$, but the proof is a little more awkward. If $p<T$, the formula in Equation (IV.37) is correct (once the denominator is modified), and applies to all the coefficients in the recursion relation (since $p-1<T-1$ and $p-1<T$). The $p=T$ case needs more finesse. There, the first sum in Equation (IV.37) is proportional to $\mathbf{C}(T-1,p-1)$, but the third sum is *not* proportional to $\mathbf{C}(T,p-1)$. That is because, for $p=T$, the lower limits of the u sums are 1, which is correct for $\mathbf{C}(T-1,p-1)$ (since $p-1=T-1$), but needs to be 0 for $\mathbf{C}(T,p-1)$. However, we can add the $u=0$ terms to *both* the third and fourth sums, and those two additions cancel (note that the limits of the t sum prevent any terms where $u=t=0$). The $u=0$ terms are not needed in the second sum, and the proof of the recursion relation then follows as with the $p<T$ case.

as above, though, the sum with a lower limit of 0, after the $u + T$ cancellation, is still 0, so the sum without that $u = 0$ term is simply the negative of the $u = 0$ term. Since the recursion relation allows us to write *any* of the $\mathfrak{d}^p$ coefficients in terms of coefficients of terms with smaller p, we can repeatedly apply it to prove that, whenever $p < T$, the coefficient is a weighted sum of coefficients with $p = 0 < T$, which we have just shown is zero. Using that, when $p = T$, the recursion relation shows that $\mathbf{C}(T, p = T) = \mathbf{C}(T = 0, p = 0)$, which can be evaluated as described previously. That proves the following remarkably simple result:

$$\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}} \left\langle \frac{u}{u+t} \right\rangle = U \left(1 - \mathfrak{t}\right)^{U-1} \mathfrak{d}^T. \tag{IV.39}$$

As a reminder, $\mathfrak{d} = 1 - \mathfrak{t}/r$ is the fraction of the trap volume that is *not* part of the interaction volume. In this form, we can trivially see that the expectation value of this slightly larger bound of $\mathcal{P}$ increases with $\mathfrak{t}$, as the derivative is a product of non-negative factors.

With the recursion relation in Equation (IV.38), we need only two simple cases to derive a formula for all of the coefficients. First, we note that if $T > 0$ and $p = 0$ in Equation (IV.35), then there is only one term in the t sum ($t = T$). Second, we lose the $\mathfrak{t}p\left(U - u\right)$ term in the brackets, and the remainder is divisible by $\left(u + T\right)$, which cancels with the $1/\left(u + t\right)$ factor (since $t = T > 0$). After these simplifications, we are left with the following sum:

$$\mathbf{C}(T > 0, p = 0) = \sum_{u=0}^{U} \binom{U}{u} \mathfrak{t}^{u-1} \left(1 - \mathfrak{t}\right)^{U-u-1} \left[u - \mathfrak{t}U\right]. \tag{IV.40}$$

Recall that we originally arrived at this by taking a derivative of a polynomial, so that the apparent singularities at $\mathfrak{t} = 0$ and $\mathfrak{t} = 1$ are not actually problems (the terms where we might have a divide-by-zero also have a zero in the numerator), and that we can safely recover the correct values by taking the limits $\mathfrak{t} \to 0$ and $\mathfrak{t} \to 1$. With this in mind, we can pull out factors of $1/\mathfrak{t}$ and $1/\left(1 - \mathfrak{t}\right)$, and this sum is essentially the mean value of $\left(u - \mathfrak{t}U\right)$ over a binomial distribution of u with mean $\mathfrak{t}U$. We could use the same techniques used to derive the mean of a binomial distribution, or we could just jump straight to the answer and state that the mean of any variable minus its mean value is always zero, which proves $\mathbf{C}(T > 0, p = 0) = 0$. The second case we care about is $\mathbf{C}(T = 0, p = 0)$. The argument is identical to the $T > 0$ case, except that, since $p = T$, the u sum starts with $u = 1$. We therefore end up with the exact same sum in Equation (IV.40), except that the u sum starts at one, not zero. In that form, though, there is no singularity for adding the $u = 0$ term, and so if we do that (and subsequently subtract that), we see that the coefficient is zero minus the $u = 0$ term, which proves that $\mathbf{C}(T = 0, p = 0) = U\,(1 - \mathfrak{t})^{U-1}$. Our two special cases are therefore:

$$\mathbf{C}(T = 0, p = 0) = U\,(1 - \mathfrak{t})^{U-1} \tag{IV.41}$$

$$\mathbf{C}(T > 0, p = 0) = 0. \tag{IV.42}$$

A quick look at the recursion relation in Equation (IV.38) shows that, for *any* $0 < p < T$, we can reduce it down to a sum of coefficients with $0 \leq p < T$, for a smaller value of p, and various values of T. Repeating this, we can reach a point where we have the sum of coefficients with $0 = p < T$ for various values of T, which, according to Equation (IV.42), is identically zero. Furthermore, if we start with $0 < p = T$, then the recursion relation states that $\mathbf{C}(T, p = T) = \mathbf{C}(T - 1, p - 1 = T - 1)$, where we took advantage of our new knowledge that, since $p - 1 < T$, $\mathbf{C}(T, p - 1) = 0$. Continuing this process, we find that $\mathbf{C}(T, p = T) = \mathbf{C}(T = 0, p = 0)$, which is given by Equation (IV.42). If we substitute these coefficients back into Equation (IV.34), then we have proved the following formula:

$$\frac{\mathrm{d}}{\mathrm{d}\mathfrak{t}}\mathcal{P} = \sum_{p=0}^{T} \mathbf{C}(T, p)\mathfrak{d}^{p} = U\,(1 - \mathfrak{t})^{U-1}\,\mathfrak{d}^{T}. \tag{IV.43}$$

As a reminder, $\mathfrak{d} = 1 - \mathfrak{t}/r$ is the fraction of the trap volume that is *not* part of the interaction volume. With the formula in Equation (IV.43), it is clear that the averaged $\mathcal{P}$ *increases* as we increase the interaction volume up to the full size of the trapping region, reaching a local maximum as the derivative is 0 when $\mathfrak{t} = r$ (and $\mathfrak{d} = 0$). This concludes our proof that our actual upper bound for $\mathcal{P}$ occurs when our interaction volume is the full trapping region. In this limit, the number of trapped atoms in the interaction region is fixed, since they are *all* in the interaction region with unity probability, and the only random variable is the number of untrapped atoms within the region. Since we will only be using upper bounds for $\mathcal{P}$, we will assume from here on that $\mathcal{P}$ is computed for the case where the interaction volume is the entire trapping volume, so that *all* of the trapped atoms are in the interaction volume, with probability unity. Therefore, we only need to average $\mathcal{P}$ over the number of *untrapped* atoms that happen to be within the trapping volume.

To place a bound on $\mathcal{P}$ given that the only free variable is now the number of untrapped atoms, we demonstrate a brief lemma. Given some function $f(u)$ that is concave down over its domain, where u is still a random variable with an expectation value of $\langle u\rangle$, we can show that $\langle f(u)\rangle \leq f(\langle u\rangle)$. The proof is rather simple. Since f is concave-down, the value of $f(u)$ is always less than the value of the tangent line.[41] We can therefore take the tangent at the mean value of u, and since this tangent is always at least as large as the actual curve, the expected value of the tangent curve is at least as large as the expected value of the actual curve:

$$\begin{aligned}\langle f(u)\rangle &\leq \langle f(\langle u\rangle) + f'(\langle u\rangle)\,(u-\langle u\rangle)\rangle\\ &= \langle f(\langle u\rangle)\rangle + f'(\langle u\rangle)\,\langle u-\langle u\rangle\rangle\\ &= f(\langle u\rangle) + f'(\langle u\rangle)\,(\langle u\rangle-\langle u\rangle)\\ &= f(\langle u\rangle)+0. \end{aligned}\tag{IV.44}$$

The remaining steps of the proof follow from the fact that the average is a linear operator (and the average of a constant is the constant value).

We apply this lemma by noticing that $u/\left(u+t+1\right)$ is concave down for all values of u (it is quite easy to take the second derivative, which is obviously negative). Therefore, since t is fixed (all the trapped atoms are within the trapping region, so

[41] In general, for a concave-down function $f(x)$, take the tangent line through x_0: $f(x_0) + f'(x_0)\,(x-x_0)$. We want to show that $f(x)$ is less than or equal to that value for all x. By subtracting $f(x_0)$ from each side of the inequality, we wish to show that $f(x)-f(x_0) \leq f'(x_0)\,(x-x_0)$ for all x in the domain. By the Mean Value Theorem, which states that the line between two points of a differentiable curve is parallel to the tangent line at some point between the two points, we know that $f(x)-f(x_0)$ is equal to $f'(x_1)\,(x-x_0)$ for some x_1 between x_0 and x. The inequality we are trying to prove then reduces to $f'(x_1)\,(x-x_0) \leq f'(x_0)\,(x-x_0)$. We can cancel the common terms (flipping the inequality if $x < x_0$), and the resulting inequality is shown to be true by applying the Mean Value Theorem to the difference $f'(x_1)-f'(x_0)$, given that $f''(x) \leq 0$ for all x, and that $x_1 - x_0$ has the same sign as $x - x_0$.

$t = T$), we know that the average maximum value of $\mathcal{P}$, $\langle u/\left(u+t+1\right)\rangle$, is bounded above by the value of $u/\left(u+t+1\right)$ when we replace u with the mean value, $\mathfrak{t}U$. We have shown that $\mathfrak{t} = r$ provides an upper bound over all sensible values of $\mathfrak{t}$, and we can also insert the value of $U = N - fN$ (from the definition of f) to get $\langle u\rangle = \left(N - fN\right) r$. Because $\mathfrak{t} = r$, the value of t is still fixed at $T = fN$, so our upper bound for $\mathcal{P}$ is:

$$\mathcal{P} \leq \frac{\left(N - fN\right) r}{\left(N - fN\right) r + fN + 1}. \tag{IV.45}$$

This is exactly what we would have used if we had not treated u as a random variable, and just assumed a perfectly smooth density.[42] Indeed, in the limit of large atom numbers (or, as described earlier, long interaction time), the fluctuations about the mean would be small enough to ignore, and this approaches the exact value of the random variable case.

The upper bound for $\mathcal{P}$ in Equation (IV.45) is only approached when the optical depth of the material is large (the large z limit), allowing for the greatest possible interaction of the pumping beam with the atoms, which in turn is achieved with the densest sample of atoms (note that in the limit of few atoms, this requires z to approach infinity, which, being much larger than the size of the atoms in the trap, and the trap itself, means this limit is infeasible in the small-atom limit). Density, here, essentially refers to the number of atoms the light interacts with over a given length of the beam, so it can be increased by increasing the cross-sectional area of the pumping beam. This limit is essentially one of efficiency. If the pumping beam is restricted to a very small volume, then the chances of it interacting with an atom are very small. In this limit, it will take many photons before we trap a single atom. Even if we know we trapped exactly one atom, the fact that it was one photon out of a large number of them implies an entropy increase proportional to the logarithm of the number of photons it took (because any of those could have caused the trapping action). Naturally, to get the lowest possible entropy increase, we can minimize that by using the fewest number of photons, which, in turn, requires maximizing the likelihood of an interaction.

[42] The same argument applies if we had used the $u/\left(u+t\right)$ upper bound for $\mathcal{P}$ instead. In that case, our upper limit is the same as this, but without the extra 1 in the denominator.

As an additional note, we note that if we use the pumping beam to trap some atoms, then, before the atoms have a chance to redisperse, we have actually *decreased* the local untrapped atom density. This *decreases* the maximum possible value of $\mathcal{P}$, so even if we do not account for suspected changes in atomic density, our value is still an upper bound for $\mathcal{P}$, which will give us a lower bound for the overall entropy change.

We now insert our limiting case for $\mathcal{P}$, given by Equation (IV.45), into Equation (IV.28) giving us a lower bound for that entropy fraction. If this value is non-negative, then the entropy change for a single photon interaction is also non-negative, and we have proved that the system entropy does *not* decrease, even though the entropy of the atoms does, eventually, decrease. To simplify, we also use the abbreviations $\Delta := (N - fN)\,r$ (note that $\Delta \geq 0$) and $x := fN + 1$:

$$\begin{aligned}
\mathcal{P} &\to \frac{(N - fN)\,r}{(N - fN)\,r + fN + 1} = \frac{\Delta}{\Delta + x} \\
\frac{\Delta S}{\mathcal{P}} &\to \{\ln(N - fN) - \ln(fN + 1) + \ln(r)\} \\
&\quad - \left(\frac{\Delta + x}{\Delta} - 1\right) \ln\left(1 - \frac{\Delta}{\Delta + x}\right) - \ln(\Delta) + \ln(\Delta + x) \\
&= -\ln(fN + 1) \\
&\quad - \frac{x}{\Delta}\{\ln(x) - \ln(\Delta + x)\} + \ln(\Delta + x) \\
&= -\ln(fN + 1) \\
&\quad + \frac{(x + \Delta)\ln(x + \Delta) - x\ln(x)}{\Delta}.
\end{aligned} \tag{IV.46}$$

We took advantage of the fact that $\ln(\Delta) = \ln(N - fN) + \ln(r)$ to cancel some of the other logarithms in the first simplification of the above expression. Note that the second term is essentially the slope between two points on the $x\ln(x)$ curve. We know $0 \leq x \leq x + \Delta$. Since $x\ln(x)$ (if we define it to be 0 at $x = 0$) is continuous over $[0, \infty)$ and continuously differentiable over $(0, \infty)$, we can apply the Mean Value Theorem. The Mean Value Theorem, roughly speaking, states that the slope between two points of a continuous, differentiable function is equal to the derivative of the function somewhere between those two points. Since the derivative of $x\ln(x)$ is $\ln(x) + 1$, that means the second term in the above equation is $\ln(c) + 1$ for some

$c \in (x, x+\Delta)$. Furthermore, the second derivative of $x\ln(x)$ is $1/x$, which is always positive over the interval $(0,\infty)$, which means that the derivative is an increasing function. Therefore, since $c \geq x$, the second term, which is equal to $\ln(c)+1$, is greater than or equal to $\ln(x)+1$. Substituting that in gives us a lower bound for our entropy change:

$$\frac{\Delta S}{\mathcal{P}} \geq -\ln(fN+1)+\ln(x)+1 = 1 > 0. \tag{IV.47}$$

The equality holds because $x = fN+1$, and we see that the entropy change in attempting to trap one more atom is positive.[43]

To sum up, we first made some assumptions that we divided the volume up into small chunks over which the atoms were evenly distributed (and implied that could always be done). We ignored many other effects (light scattering, atomic recoil, and so on) that should serve to increase the entropy change of the system. We kept only those effects that were absolutely necessary for the barrier to work in an asymmetric (one-way) manner. This means that we should be computing a lower bound for the actual entropy change. With these assumptions, we wrote up expressions for the entropy of the atoms and the pumping beam through the interaction with the atoms. We then computed bounds on these entropies assuming we shone just enough pumping beam light to possibly trap one (but not two) atoms. We performed some

[43] If we were using the $u/(u+t)$ upper bound for $\mathcal{P}$, the result is the same, except with $x = fN$. In this case, our lower bound simplifies to $1-\ln(1+1/x)$, which is nonnegative if $x = fN \geq 1$ (at least one atom is already trapped), but becomes negative as x approaches 0 (the case where no atoms are trapped). The problem is that in the limit where no atoms are trapped, our bound for $\mathcal{P}$ becomes 1, meaning an atom will be trapped with certainty, and so there is no increase in photon entropy to counter the decrease in entropy due to trapping an atom (unless the atoms are distinguishable, in which case there is a countering term from which atom became trapped). We dealt with this case by appealing to the discreteness of atoms, and stating that if $fN < 1$, it had to be 0 (no trapped atoms). In the fast-interaction limit (the $u/(u+t)$ limit is too weak to work with the long-interaction limit) there is still a nonzero probability that there are no atoms in the trapping volume, in which case no atoms can be trapped. That yields $\mathcal{P} \leq 1-(1-r)^{N-fN}$. With $N=2$, $fN=1$, this gives a negative lower bound for entropy change, but gives a non-negative bound for the only case we need, $f=0$. The bound may be written $f(1)-f(1-r)$ with $f(x)=x^N$. By the Mean Value Theorem, that is $rf'(c)$ for some $c \in (1-r,1)$. $f(1)-f(1-r)$, where $f(x)=x^N$. The Mean Value Theorem states that this is equal to $rf'(c)$, for some c satisfying $1-r < c < 1$. Since f' is an increasing (or constant, if $N=0$ or $N=1$), we have $\mathcal{P} \leq f(1)f(1-r) \leq rf'(1) = rN$. Substituting $f=0$ and $\mathcal{P} \to rN$ into Equation (IV.28), using $-(1-\mathcal{P})\ln(1-\mathcal{P})/\mathcal{P} \geq 0$ since $0 \leq \mathcal{P} \leq 1$ (it is a probability), and the other bound $\mathcal{P} \leq rN$ for the other $\mathcal{P}$ term, yields a non-negative bound for the entropy change.

algebra on the expression, and reduced it to a quantity which has the same sign as the overall entropy change, and then demonstrated that this term is always non-negative. This proves that with each trapping opportunity, the entropy of the system increases. Therefore, the overall change over many such events, even if we need to remap the volumes in between, must be an overall increase. Even though our one-way barrier appears to violate the second law of thermodynamics, we have shown that when the entire system is accounted for, the pumping beam carries away some information about the initial distribution of the atoms. Because the trapping is not deterministic, the number of possible ways this information can be represented in the pumping beam is larger than the number of distributions of the atoms, and the net result is the entropy of the entire system increases, even though the atoms alone end up in a state with less entropy than the initial configuration.

In the language we used to discuss why a one-way barrier such as this cannot be used as a perpetual motion machine, we have a system (the atoms) where we can decrease the entropy. The reason this can happen is that we have a low-entropy memory, which is the pumping laser beam, consisting of photons that are all the same. Since we have a low-entropy reservoir, we do not need to directly perform work to reduce the entropy of the atoms by increasing the entropy of the pumping beam. However, to produce the laser beam (or the erased memory) in the first place, we must either have a different, lower-entropy source, or we must perform work to "pump" entropy out of the beam we are creating into the high-entropy environment. As a result, the resolution of our apparent Maxwell's demon paradox is either that we do need to do work somewhere to create a low-entropy beam, or we can simply state that we start with a low-entropy beam. Since it takes very little energy to increase the entropy of that beam (technically, because the photons emitted when we trap an atom are of a lower frequency, we actually get energy out of the process), the temperature of that beam, which is proportional to the derivative of the energy as a function of entropy, is essentially zero. In this case, though, the entropy is not really a function of the energy of the beam, or vice versa, so this description involves a little hand-waving, but helps to get the point across. If we choose to state that we start with a low-entropy beam, then we basically have a low-temperature reservoir into which we can dump entropy, and the paradox is resolved: Entropy is not decreasing, we are just transferring it from a high-entropy source to a low-entropy reservoir, rather like letting energy flow from a high temperature system to a low temperature reservoir. If we were to try to sustain the process, we would need to continue creating a low-entropy reservoir, which, at some point in the process, requires more energy than we could possibly extract from our one-way barrier scheme. Even without that requiring energy transfers, though, we have explicitly shown that entropy cannot decrease in our one-way barrier, so the second law of thermodynamics is upheld.

CHAPTER V

ELECTRON-MULTIPLYING CHARGE-COUPLED DEVICES

We have finished what we planned to do with the one-way barrier experiment, and are now moving on to what the experimental setup was intended to probe: The transition between classical and quantum mechanics. By closing off the rubidium source in the vacuum chamber and waiting for months, we have reduced the rubidium pressure in the chamber so that the MOT loads very slowly. It is now feasible to load a single atom into the dipole trap by loading the MOT for a very short period of time before loading the dipole trap.

Current theories of quantum mechanics predict that classical mechanics emerges from quantum mechanics through a coupling to an external environment, such as via a measurement [29]. While quantum mechanics and classical mechanics can be compared by comparing probability distributions, using a measurement provides a measurement record that can be directly compared with classical trajectories. Without going too deep into the theory of measurements, we can say that a measurement tends to localize quantum mechanical wave packets. If the measurement indicates that an atom is probably in a certain location, we can update our idea of the atom's wavefunction by localizing it in that location. In doing so, we prevent the packet from spreading out through space, reflecting off the walls of some confining potential, and interfering with itself. If the localization is strong enough, but not strong enough to scatter the atom away, the wave function acts like a point particle moving according to classical rules [28].

An atom in a dipole trap is ideal for probing such a transition. The atom is neutral, and has a weak magnetic moment, and so couples very weakly to external static electric and magnetic fields (and we can cancel magnetic fields with our Helmholtz coils). The vacuum is clean enough that there is very little within the chamber for the atom to interact with. Indeed, the only thing the atom interacts strongly with is light that we provide. The dipole trap itself is far enough off-resonance that it should not disturb the atom other than create a conservative potential in which the atom moves. We can then also illuminate the atom with a weak pulse of resonant light, which will scatter off the atom and, if measured, provide a position measurement of the atom. The strength of that field controls the strength of the measurement. In the limit of no scattering, the atom should behave according to the rules of quantum mechanics. As we increase the intensity of the resonant light, we should be able to measure some statistics of the atom's motion. When the measurement becomes strong enough that we can localize the atom faster than the wave packet spreads through the trap, we should see the atomic statistics change to a more classical version [28].[1] Current theories also allow for interesting results strictly from the measurements themselves [30, 31].

The largest experimental barrier to this problem is the measurement. Seeing the quantum-mechanical behavior requires collecting statistics at the near-single-photon limit. In order to get some idea of the motion, we would need to image the atom multiple times for each period within the trap. The trap period would probably be on the order of tens to hundreds of milliseconds, so we would need to be able to acquire images on the order of tens to hundreds of frames per second. Furthermore, in order for the measurement to be weak enough, the resonant light would need to be weak enough to only scatter a few photons during the trap period, so each frame would only collect a few photons from the atom. While there are photon-detecting devices that could detect the photons with the efficiencies we would need, they are expensive, and using a large array of them for imaging would be difficult. However, there is now a type of camera called an electron-multiplying charge-coupled device (EMCCD) camera that is capable of the frame rate and spatial resolution we think we would need, and these cameras have almost single-photon detection capabilities.

In particular, these cameras have pixels that are on the order of ten microns on a

[1] Quantum mechanics and classical mechanics actually predict the same statistics for free particles and particles in harmonic potentials. In order to see different statistics, we would need to have a more complicated potential. The most interesting potentials would be those which have chaotic classical trajectories, such as a driven double-well potential. By scanning off-resonant light across the dipole trap and varying the potential in a position-dependent manner, we should be able to "draw" arbitrary potentials in which to confine the atom.

side, about 70% quantum efficiency at the near infrared wavelengths we use, and can almost count single photons. We feel that EMCCD cameras could be used for such an experiment, provided we develop a thorough understanding of how they work, what their noises and inefficiencies are, and why we cannot quite count photons with them.

This chapter presents the basic theory of how an electron-multiplying charge-coupled device (EMCCD) works. We also present an in-depth description of the noise statistics for EMCCDs, and show how single photons are pretty well differentiated from no photons (although there is some overlap), but higher numbers of photons overlap quite a bit.

Charge-Coupled Device (CCD) Operation

We do not intend for this to be a full description of how a charge-coupled device (CCD) works. This is intended to be a short description of how CCDs function, to allow us to develop a model of the noise sources in a CCD [38, 39].

There are several types of CCDs, including *full-frame*, *frame-transfer*, and *interline* CCDs. These all work off the same principles, and differ only in their geometry. A CCD is a semiconductor device that converts light to charge. Specifically, when light hits the CCD, it promotes electrons to the conduction band of the semiconductor. For the purposes of this chapter, we will assume that a single photon of incident light energy promotes exactly one electron to the conduction band of the semiconductor, with probability Q_{eff}. Q_{eff} is the *quantum efficiency* of the CCD, and gives the fraction of incident photons that are converted to free charge in the CCD:

$$N_{\text{el}} = Q_{\text{eff}} N_{\text{ph}}. \tag{V.1}$$

where N_{el} is the number of electrons promoted to the conduction band, and N_{ph} is the number of incident photons. Attached to the light-sensitive region of the CCD are an array of electrodes called the *gate electrodes.* Voltages applied to these electrodes are used to create potential wells for the conduction electrons, restricting them to square cells on the CCD. These square cells form the pixels in the final image taken with a CCD camera.

At the beginning of an exposure, all the electrons are ideally in the valence band. During the exposure, light hits the various cells, promoting electrons to the conduction band. These electrons are physically restricted to their cell of origin by the electric potentials of the gate electrodes. At the end of the exposure, an image is formed by counting the electrons in each cell. Each pixel of the image is given an intensity (or color) based on the number of electrons found in the corresponding cell on the CCD. Rather than actually count the electrons, CCDs use a *readout amplifier*, which is essentially measures the voltage across a cell from the electrons in the conduction band, and convert that voltage to a number. Since the voltage is very closely proportional to the number of electrons, this voltage measurement represents the number of electrons in the cell.

There is a type of image sensor, called a *CMOS sensor*, that has a readout amplifier for each cell. CCDs employ a method for sharing a readout amplifier, which is what allows for the extended amplification register used in EMCCDs. This also allows for the CCD manufacturers to use a single (or perhaps just a few) very sensitive and linear readout amplifier across a large segment of the image, reducing artifacts from non-uniform amplifiers. The readout amplifier only reads one cell, and the gate electrodes are used to move the charge to the readout amplifier. With enough gate electrodes, the electrode voltages can be varied in such a way as to smoothly move the cells across the CCD to carry the charge to the readout amplifier. We show a simplified, one-dimensional version of this in Figure 5.1.

There are three common types of CCD that differ mainly in the path the charge is moved through to get to the readout amplifier [38]:

1. The *Full Frame CCD*

2. The *Frame Transfer CCD*

3. The *Interline CCD*

The full frame CCD is the simplest, with the readout amplifier next to the lower-

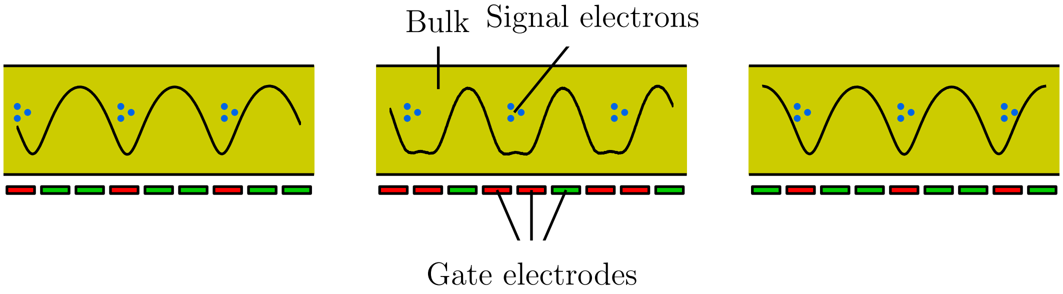


Figure 5.1. A simplified schematic of the clock cycle used to shift electrons around a CCD. Electrons, excited into the conduction band of the bulk by illumination, are restricted to cells by electric potentials applied to the gate electrodes. Shifting the potentials applied to the electrodes in sequence moves the charge, and hence the recorded light signal, around the chip. We show three such steps in sequence from left to right. The electrodes are colored by whether or not they are activated.

right corner of the CCD cells (we're using the location of the readout amplifier and the rows and columns of cells to define a coordinate system). After the exposure is complete, the cells of the bottom row are shifted to the right one cell. This puts the charge that was in the lower-right-most cell of the CCD into the cell that the readout amplifier reads. The readout amplifier measures the charge in the cell, and the camera electronics record that value in the lower-right corner of the image. The bottom row is then shifted to the right again, and the next cell's charge is recorded, and placed in the appropriate part of the image. Once an entire row is read by the readout amplifier, the gate electrodes are used to shift each entire row of the CCD down one full row. The bottom row will then contain the charge that was formerly in the second to last row of the CCD, and the horizontal shifting and readout can begin for that row.

The entire image in a full-frame CCD is read out this way, shown in Figure 5.2: The bottom row is shifted to the right, placing each cell from that row into the readout cell, where it is read. Once an entire row is read, each row is shifted down, so that the next row may be read. This method, while simple, has a flaw: Readout amplifiers and cell shifting can only be done so quickly. The faster a readout amplifier reads, the more noise it will produce. For the EMCCD cameras we tested in our lab, the read-rates for the readout amplifiers typically varied from about 100 kHz to 10 MHz [101, 102]. For a CCD with a 512-by-512 array of cells, that means that it would take approximately 25 ms to read out every cell in the array, even at the fastest speed. If we were imaging even a moderately bright source of light that could saturate a CCD cell in that period of time, then, as the cells get shifted, that light would partially expose every row above the row the light was originally incident on, creating a streak in the final image.

A frame-transfer CCD greatly reduces this smearing problem at the expense of the CCD needing to be twice as large as it would otherwise be. In a frame-transfer CCD, the lower half of the CCD is masked, meaning it is covered by something that prevents any light from hitting it. At the end of the exposure, rather than horizontally shifting the bottom row, the entire CCD is immediately shifted down, row by row, so that the entire exposed region is shifted under the masked region. No readout is done during this vertical shifting, which can be very fast. This, combined with the fact the number of shifts needed is only the number of rows (instead of the total number of cells), gives much less time for excess exposure. We show this readout process in Figure 5.3. One camera we tested advertised a vertical shift rate of 200 MHz, so that 512 rows can be shifted to the masked region in about 2.5 μs [102]. An extension of the frame-transfer CCD is the interline CCD, where instead of having the sensor split into an unmasked half and a masked half, every other row is masked. This eliminates smear by only requiring one vertical shift to move the charge under a mask. Once shifted, the individual rows may be shifted to a readout register, and then read [38]. The downside to an interline CCD is that half of the exposed sensor is masked, reducing the amount of light collected. This can be reduced with *microlenses*, which are tiny lenses on the chip itself that focus light onto the unmasked regions, but interline CCDs typically do not have the quantum efficiency that frame-transfer CCDs have [38].

One more image artifact that happens with CCDs is called *blooming*. Blooming is where one cell (or more) of the CCD collects its maximum amount of charge. With sufficient charge, the repulsion of the electrons overcomes the potentials that normally confine the electrons, and they leak into neighboring cells. Blooming tends

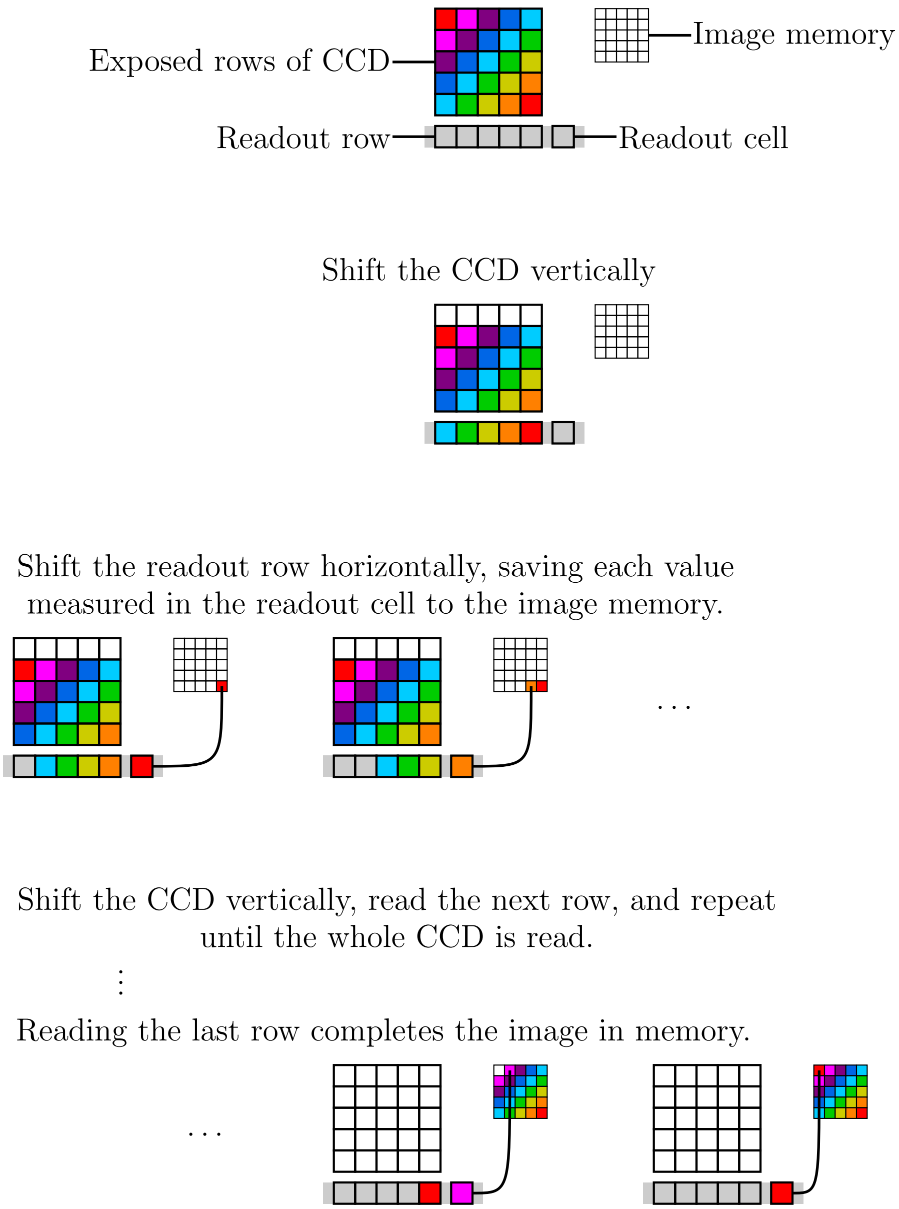
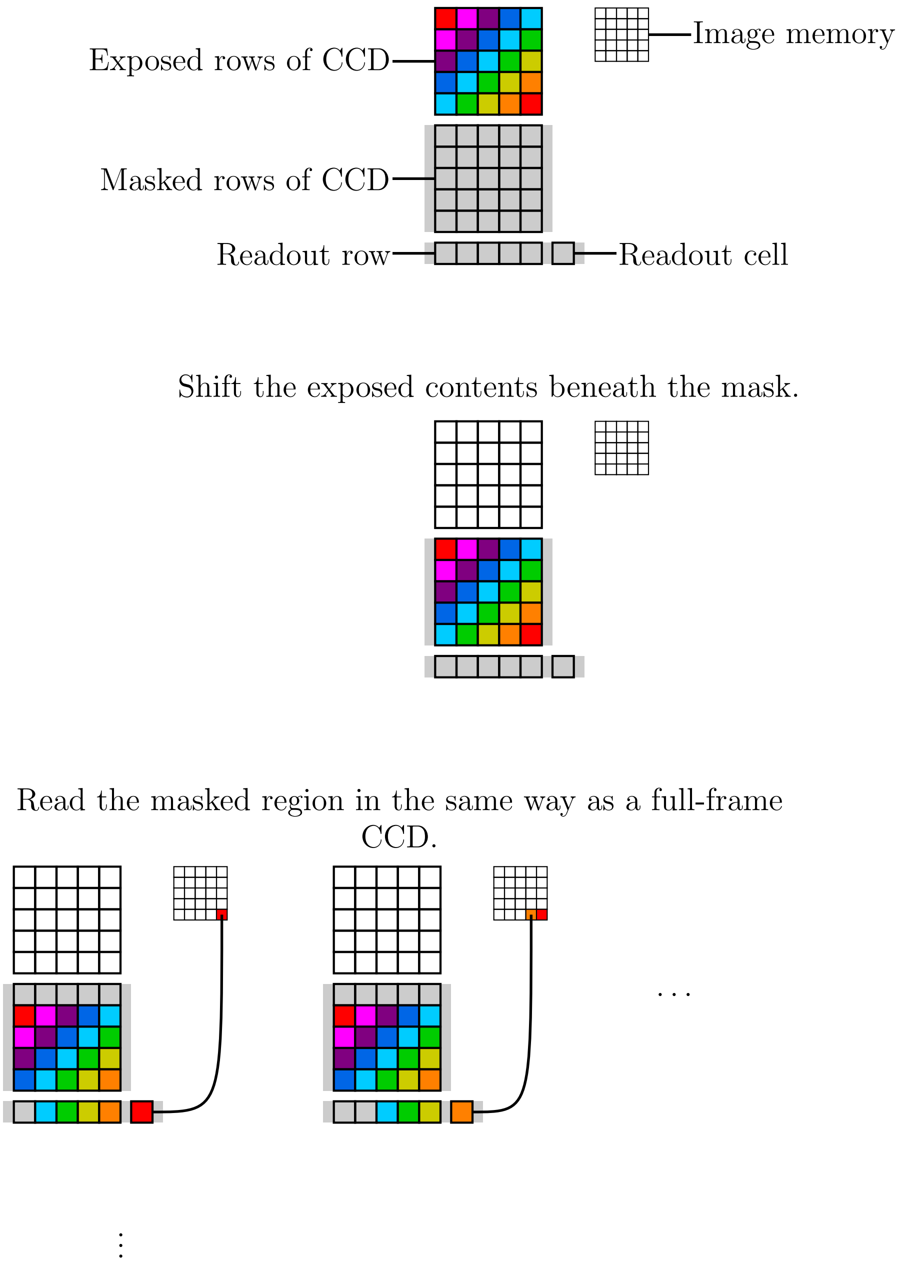

Exposed rows of CCD
Image memory
Readout row
Readout cell
Shift the CCD vertically
Shift the readout row horizontally, saving each value measured in the readout cell to the image memory.
...
Shift the CCD vertically, read the next row, and repeat until the whole CCD is read.
⋮
Reading the last row completes the image in memory.
...

Figure 5.2. A schematic representation of a full-frame CCD in operation. A region of the CCD is exposed to light in the same way as a piece of film in a camera, forming an image on the CCD. During the exposure, the CCD records the total amount of incident light on each cell as the number of electrons in the conduction band of that cell. The number of electrons stored in each cell is represented here by an arbitrary color that serves to show the initial location of that charge. Through a series of shifts, charge is moved to the readout cell, where it is measured and recorded, building up an image in memory.



Shift the exposed contents beneath the mask.

Read the masked region in the same way as a full-frame CCD.

Figure 5.3. A schematic representation of a frame-transfer CCD in operation. The process is the same as for the full-frame CCD shown in Figure 5.2, except between exposure and readout, during which the charge in the entire frame is shifted down to a masked region that is protected from external light.

to occur along the image axes, creating a somewhat star-shaped pattern in a bright region of the CCD.

Description of Noise Sources in Charge-Coupled Devices (CCDs)

As with all sensitive devices, the readout of a CCD contains noise. The noise on a CCD can be characterized by a probability distribution that gives the probability of getting a certain output value for a given cell. In this chapter, that distribution will be called a *cell-charge distribution.* The cell-charge distribution can be measured by reading the exact same image out many times, and computing a histogram of the values for a given cell. For the case of a dark frame, where the entire CCD has not been exposed to any light, and ignoring any position-dependence to the noise, the cell-charge distribution also gives the histogram for a single image, or a conglomerate of images.

There are three main sources of noise in a CCD:

1. *Thermal noise*

2. *Clock-induced charge*

3. *Readout noise*

Thermal noise is also called *dark current* [38].

Thermal noise is the spontaneous excitation of valence electrons into the conduction band. The rate of this excitation is dependent on the temperature of the CCD substrate, and is approximately constant, independent of the amount of charge already excited in the cell [41]. As such, this produces a Poissonian cell-charge distribution, described in Section B.4. As the process happens at a constant rate, the mean and variance of this distribution increase linearly in time. Thus, the thermal noise is often quoted as a dark current on camera specification sheets, which gives an order-of-magnitude for the rate at which the mean and variance of the thermal noise increase over time [38, 101, 102].

Thermal noise is commonly reduced by simply reducing the temperature of the CCD substrate [41]. Scientific-imaging CCD cameras typically employ thermoelectric coolers backed by air-cooled or water-cooled radiators to cool the CCD down to below -50°C. The cameras we had as demonstration units cooled down to between -65°C and -75°C using room-temperature air-cooling only. At these temperatures, the thermal noise (called dark current on many camera specification sheets) is reduced to on the order of 0.01 to 0.001 e^-/pixel/s (electrons per pixel per second) [38, 101, 102]. For shorter exposures (less than a tenth of a second), we will see that this value is typically small compared to the clock-induced charge, and hence almost negligible.

Clock-induced charge (CIC) is the excitation of valence electrons into the conduction band by the voltages from the gate electrodes while moving charge from one cell to another [38, 101, 102]. Supposedly, no charge is lost as it is moved from one cell to another (or, at least, the loss is very small), but some excess charge is gained. This excess charge is called the clock-induced charge [41]. The clock-induced charge from a single shift is very small, and supposedly roughly independent from the amount of charge in that cell. Thus, the cell-charge distribution for the CIC from one shift is closely approximated by a Poissonian distribution, as described in Section B.4. A second (and a third, and so on) shift adds another charge with a Poissonian distribution, and the mean is independent of the amount of both the charge in the cell and how much of that was from the previous shift. As demonstrated in Section B.4, the sum of independent Poissonian distributions is another Poissonian distribution, with the means adding. This means that the cell-charge distribution for N identical shifts is a Poissonian distribution, with a mean of N times the mean for a single shift.

Although one might expect the CIC to be larger in the upper rows of an image, since they have to traverse more rows to get to a readout register, it is, in fact, almost constant across an image. This is because the CCD is initially cleared by vertically clocking the charge from the top to the bottom. Thus, while the upper row has to be moved across the entire N rows on a CCD after the exposure, and the bottom row only needs to be moved down one row to the readout row, before the exposure, the upper rows was clocked one row from the region above the exposed area, and the bottom row was clocked N rows down from that same region. Thus, both rows have experienced $N+1$ vertical shifts by the time they reach the readout row. CIC happens mostly during the vertical shift, where the charge in each row is moved one row down, as opposed to a horizontal shift, where charge is moved across a readout row, and so little change in CIC is expected across a single row of an image [41]. For the 512-row frame-transfer cameras we tested (since frame-transfer cameras have twice as many rows as are available for exposure, that is 1024 rows plus a readout row), the CIC specification was not always given. When it was given, the number given was typically 0.001 to 0.01 e^{-}/pixel [38, 102]. We mistakenly assumed that this number was the mean value of the Poisson distribution for CIC, but this turned out to be incorrect. We will defer what these numbers actually mean until Section V.7, when we will have the formulas available to properly relate them to the actual mean.

Readout noise refers to any added noise from the measurement of the amount of charge in a cell and the conversion to a digital number. This is typically the sum of many complicated electronic processes, with some independence, and so, by the Central Limit Theorem discussed in Section B.3, we expect this to be well-approximated by a (discrete) Gaussian distribution. Looking at actual histograms from images confirms that this distribution is, to a very good approximation, Gaussian, with the deviations out in the tails where the distribution is very small. Technically, this is not a cell-charge distribution, as the noise is not actually present in the charge distribution, but is added either during or after measuring the charge. However, it can be modeled as a Gaussian noise added to the cell-charge distribution and then measured in a noise-less fashion, and so we will treat it as an effective part of the cell-charge distribution. The cameras we looked at quoted readout noises as the standard deviation in the effective Gaussian cell-charge distribution. The values were dependent on the readout speed, and were typically in the range of 2 to 50 electrons. Faster readout tend to have more noise, and various other tricks (like having multiple readout cells) can effect it as well. As described in Section V.3, for EMCCDs there is an effective readout noise that is less than the actual readout noise, and this number is sometimes quoted instead of the actual readout noise in camera specifications [101, 102].

There are a few other parameters CCD manufacturers have to help with noise. In particular, the exact waveform applied to the gate electrodes can be used to reduce both thermal noise and clock-induced charge. Perhaps a better way to state this is to say that a bad waveform will greatly *increase* these noise sources. The clock waveforms are a pretty delicate science. For example, thermal noise occurs anywhere within the bulk of the CCD, but clock-induced charge tends to occur mostly on the surface of the CCD. The polarity of the clock waveform can actually selectively inhibit either thermal noise or clock-induced charge. For short exposures, thermal noise is often negligible next to clock-induced charge. For long exposures, thermal noise is often dominant. Thus, EMCCD camera manufacturers often use multiple waveforms, depending on the type of imaging being done. In the cameras we investigated, the optimization is often chosen based on the readout rates selected by the user. Faster readout rates are assumed to be coupled with shorter exposures, where thermal noise is less important, and so they tend to optimize for less clock-induced charge, and slightly more thermal noise. Slower readout rates are assumed to be for longer exposures, and so they tend to optimize for less thermal noise at the expense of slightly greater clock-induced charge [38].

We were searching for a very high quantum efficiency with low noise and high gain, and all the candidates we looked at were back-illuminated frame-transfer EMCCDs. Thus, for the remainder of this chapter, we will deal only with this type of EMCCD.

Electron-Multiplying Charge-Coupled Device (CCD) Operation

Electron-Multiplying Charge-Coupled Devices (EMCCDs) are very similar to CCDs. The only difference is that they have an electron-multiplying stage in the readout row to amplify the signal before it gets measured. This reduces the effect of the readout noise [38]. To see how, assume a model where you have a signal you want to measure, and you add thermal noise, and CIC, and the readout stage adds a Gaussian noise to the signal, which is then measured by an ideal noise-less amplifier. The signal which reaches the amplifier is (in number of electrons):

$$\text{final signal} = \text{signal} + \text{thermal noise} + \text{CIC} + \text{readout noise}.$$

All the noise sources are independent of each other and the signal and have a fixed mean, which can be subtracted away to reveal the signal. Thus, it is only the deviation from the mean of the noise that matter. As mentioned in Section B.1, the variances of the independent noise sources add. The total noise variance (in number of electrons) is therefore:

$$\text{var(noise)} = I_{\text{dark}}t + \sigma_{\text{CIC}} + \sigma_{\text{r}}^2 \tag{V.2}$$

where $I_{\text{dark}}t$ is the dark current times the exposure time (which, for Poisson noise, is both the mean and variance in the number of electrons), and σ_{CIC} and σ_{r}^2 are the CIC and readout noise variances, respectively. The signal to noise ratio (SNR) is the mean signal value divided by the deviation of the noise:

$$\text{SNR} = \frac{\text{signal}}{\sqrt{I_{\text{dark}}t + \sigma_{\text{CIC}} + \sigma_{\text{r}}^2}} \tag{V.3}$$

Unless the exposure is very long, σ_{r}^2 is the only term in the denominator that is larger than 1 electron squared, so it is the dominant term. EMCCDs make σ_{r}^2 effectively smaller by amplifying everything before the readout stage by some gain G. This process, as we will see later, introduces some extra noise, so the variances get multiplied by fG^2, where f is some extra noise factor that approaches $\sqrt{2} \approx 1.4$. With this gain, the SNR becomes:

$$\text{SNR} = \frac{G \times \text{signal}}{\sqrt{fG^2 I_{\text{dark}}t + fG^2\sigma_{\text{CIC}} + \sigma_{\text{r}}^2}} = \frac{\text{signal}}{\sqrt{fI_{\text{dark}}t + f\sigma_{\text{CIC}} + \sigma_{\text{r}}^2/G^2}} \tag{V.4}$$

This is essentially the same as Equation (V.3), but with an effective readout noise variance of σ_{r}^2/G^2. The deviation associated with this (σ_{r}/G) is the effective readout noise sometimes quoted in camera specifications. Since σ_{r} is typically less than 100 electrons, and EMCCD gains typically have $G \sim 1000$, the effective readout noise is almost always less than 1.

Note that Equation (V.4) has a reduced effective readout noise, but the thermal and CIC noise values remain effectively unchanged. This is because thermal noise and CIC result in electrons being escalated to the conduction band in the CCD, just like signal. They are effectively indistinguishable from the signal, and get amplified just like the signal.

The actual gain is produced in a manner similar to photomultipliers. Some extra cells (~ 500) are placed in the readout row, before the readout register, as shown in Figure 5.4. As charge is horizontally shifted through these each of these extra cells, there is a large voltage present from one cell to the next (large, for a CCD, means ~ 30 V). This voltage accelerates electrons as they pass from one cell to the next, such that the electrons can collide with valence electrons and move them to the conduction band as well, with some small probability g. Thus, with each step, the number of electrons is increased by a fraction g of the current number of electrons. This fraction is typically a few percent, but after $N \sim 100$ cells, the total gain $G = (1+g)^N$ can be over a thousand [38].

Because the signal and noise are independent, we can write out the cell-charge distribution entering the EM stage. The amount of charge in the cell is the sum of two independent variables (the signal and noise), and so, as shown in Section B.2, the cell-charge distribution is the convolution of the distributions of the signal and the noise:

$$\mathcal{H} = \mathcal{H}\{\text{signal}\} * \mathcal{H}\{\text{noise}\} \quad \text{(before EM stage).} \tag{V.5}$$

The noise cell-charge distribution is the Poissonian distribution given in Equation (B.23). The mean and variance of the charge in the cell before the EM stage is just the sum and variance of the two distributions:

$$\text{mean of cell charge} = (\text{mean of signal}) + \sigma_{\text{CIC}} \tag{V.6}$$

$$\text{variance of cell charge} = (\text{variance of signal}) + \sigma_{\text{CIC}}. \tag{V.7}$$

Here, we have used the fact that the mean and variance of the noise are both σ_{CIC}.

As the charge in the cell passes through the EM stage, more charge is added to the cell. This charge has the effect of multiplying the amount of charge in the

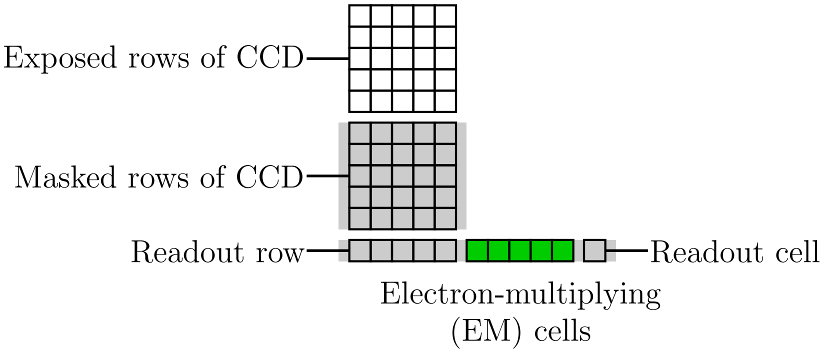


Figure 5.4. A schematic representation of a frame-transfer EMCCD. It is basically identical to a frame-transfer CCD, with an exposure area and a masked copy to store the charge from the exposure during readout. The only difference is the readout row has some extra cells which typically have larger voltages across them, which multiply charges as they traverse this region before they reach the readout cell.

cell by some gain G, which means the charge is *not* independent of the amount of charge already in the cell. Therefore, we cannot simply add means and variances together, but we can compute the mean and variance of the charge in the cell after the EM stage. We will do this shortly.

Once through the EM stage, the readout cell adds a small amount of Gaussian noise to the charge in the cell. Although the noise should have zero mean, many of the amplifiers add a constant offset to the signal before converting it to a digital number. This is because they usually convert to a non-negative integer, which would clip the readout-noise Gaussian for zero input charge, which would alter the average signal of an image. Whether this offset is done by actually adding a voltage to the signal or by just manipulating the digital number afterwards, it can be modeled by assuming the readout-noise has a non-zero mean. In any case, the readout noise is again independent of the amount of the charge in the cell, and so the final cell-charge distribution is the convolution of the post-EM cell-charge distribution with a Gaussian of some arbitrary mean and standard deviation σ_{r}. Likewise, the mean and variance are the post-EM mean and variance added to the readout-noise mean and variance (σ_{r}^2), respectively.

Comparing EMCCD Dark-Frame Histograms

Naturally, we wish to have some way to quantitatively compare cameras, and, for theoretical use, we would like some quantitative model of how EMCCDs work. Since we intend to use these EMCCD cameras to image very faint objects, we want to focus on the noise sources and quantum efficiency of EMCCDs. The easiest way to measure intrinsic EMCCD noise is to take an image with no signal whatsoever, in what is called a dark frame. Taking dark frames is easy. All the EMCCD cameras we looked at came with a cap that blocked pretty much all light that could reach the sensor (and some had an extra shutter or two as well). We simply turned out the lights in the room to make it fairly dark, placed the cap on, closed any available shutters, and tried taking images at the most sensitive setting. Typically, the average pixel value of the images did not change significantly with the lights on versus with the lights off, so we figure that with the lights off, the amount of external light leaking onto the camera sensor was negligible.

To model the noise in the camera, we chose to look at camera histograms, which are probability distributions of each possible value for the pixels. We start by making some fairly justified assumptions. First, at full EM gain, the EM gain stage typically has gains of about 1000, with readout noise greater than 10 e^{-}, and the CIC has a mean of less than 10%. Furthermore, the EM gain stage may have over 500 separate steps. These allow us to make assumptions that the EM gain and readout noise are large compared to the CIC, and that the EM stage is essentially continuous. In Appendix A, we justify these assumptions a little more, and use them to derive a formula for what an EMCCD dark-frame histogram should look like [38, 101, 102]. The formula we use to fit dark-frame histograms (and also work for very low light levels) is Equation (A.56), which we reproduce here (with more variables defined):

$$\mathcal{H}\{\text{final}\}(x) \approx \left\{ \frac{e^{-\dfrac{x^2}{2\sigma_{\mathrm{r}}^2}}}{\sqrt{2\pi\sigma_{\mathrm{r}}^2}} \right\} * \left\{ \sum_{k=0}^{\infty} \frac{\Gamma(x+1+L_G)G^{-L_G-1}}{\Gamma(k+L_G)\Gamma(x+1)} \left(\frac{x}{G}\right)^{k-1} e^{-\frac{x}{G}} \mathcal{H}\{N_0\}(k) \right\}$$

$$\mathcal{H}\{N_0\}(k) = \frac{\sigma_{\mathrm{CIC}}^k}{k!} e^{-\sigma_{\mathrm{CIC}}}$$

where the $k = 0$ term of the sum approaches a delta function in x as $L_G \to 0$. $\mathcal{H}\{N_0\}(k)$ is the pre-EM stage histogram, giving the probability of having k electrons in any given pixel before entering the EM stage. This is assumed to be a Poissonian distribution with mean σ_{CIC} (shown above). We assume we will be dealing with short exposures where the thermal noise is negligible, but since the sum of independent Poissonian noises are also a Poissonian noise (as shown in Section B.4), it is correct to just assume σ_{CIC} is the sum of the mean CIC noise, the mean thermal noise, and the mean input photon value.[2] $\mathcal{H}\{\text{final}\}(x)$ is the final histogram, either in electrons or *Analog Data Units* (ADUs), which is the units output by the camera, and are proportional to the number of electrons, but the proportionality does not need to be known (often, there is an offset added on as well, in which case x in this formula should be shifted by that offset). The other parameters are:

1. σ_{CIC}, the combined CIC and thermal noise (and maybe low light levels), in number of electrons

2. σ_{r}, the readout noise (in electrons or ADU)

3. G, the EM stage gain (in electrons or ADU)

4. L_G, the EM leakage rate divided by the natural log of G

5. the offset applied by the readout stage (in ADU)

[2] This is zero for a dark frame, but since low-level light sources tend to have Poissonian statistics, we can include that here as well.

Naturally, if one parameter is in ADU, then all the parameters with that option must be scaled to ADU as well. As mentioned, the readout stage offset is not explicitly shown in the equation (we have effectively set it to zero, but it just translates the histogram). The EM leakage is basically CIC noise that happens as charge is moved through the EM stage. The EMCCD engineers we communicated with did not seem to know of any such effect, but we will show shortly that this model performs decently without the leakage term, and exceptionally well with the leakage term.

We feel we should mention that our EMCCD model also applies to certain photomultiplier models where the electron multiplication can also be approximated as a continuous random gain stage. To have our EMCCD model match these particular photomultiplier models, we need to take into account that the electron cascade is amplified enough that reaodut noise is negligible, and assume that the multiplication happens nearly instantaneously. The latter assumption means that we only need to work out the distribution of a single input electron, and we can ignore leakage. Our distribution with these assumptions, for any number of input electrons, is given by Equation (A.42). For the particular case of exactly one input electron, it reproduces one of the standard output distributions for photomultipliers [103, 104].

Equation (A.56) (reproduced above) looks very complicated, but there are some simplifications worth noting. These are all derived in Appendix A, so we will only mention them here. The terms inside the sum reduce to Poisson distributions in the $L_G \to 0$ limit. They represent the histogram that results when exactly k electrons pass through the EM stage. We then weight this by a Poisson distribution ($\mathcal{H}\{N_0\}(k)$) representing the probability of having k enter the EM stage, and sum over that. Then, we convolve with a Gaussian to represent the addition of noise in the readout stage. We used several approximations to get the form of the equation such that, as long as the readout noise and EM gain are both scaled by the electrons/ADU conversion, it still gives (to a pretty high degree of precision) the correct histogram even after the readout stage scaling (and the shift is simply a translation in x). Because we can scale x without changing the equation in this approximation, we can always scale x so that the readout noise is 1, removing one of the parameters. In this way, we see that the actual parameter is the ratio of the EM gain G to the readout noise σ_r, which is independent of the readout stage scaling.

Since $\mathcal{H}\{N_0\}(k)$ is effectively bounded in a dark frame (we are assuming it is a Poissonian distribution with a very low mean), the exponential in the sum in Equation (A.56) dominates in the large-x limit, so having a non-zero CIC results in an exponential-like tail. In a dark frame with low CIC, the dominant feature is a delta-function-like peak corresponding to most of the pixels having 0 electrons in them; once convolved with the Gaussian readout noise, this results in the entire histogram looking like a large Gaussian with a near-exponential tail. In Appendix A, we show this explicitly for the small-CIC and small-leakage limit by actually computing the first two terms, which look like Equation (A.52), a Gaussian, plus Equation (A.53), which, for $x \gg \sigma_\mathrm{r}$, is almost exactly exponential. See Figure 5.5 for a graphical demonstration of the parts of Equation (A.56), and to see how a real histogram ends up resembling a Gaussian plus an exponential tail. In the figure, we start with a Poissonian distribution representing a typical CIC value, although it could also represent a low amount of either thermal noise or dim light sources. This is given by $\mathcal{H}\{N_0\}(k)$ in Equation (A.56). After passing through the EM stage, cells with 0 electrons are unaffected (the near-delta-function behavior of the $k = 0$ term of the sum in Equation (A.56)), while cells with multiple electrons have those electrons multiplied by roughly a thousand-fold (the Poisson-like distributions of the $k > 0$ terms of the sum in Equation (A.56)). Finally, when passing through the readout stage, the histograms are convolved with Gaussian readout noise, producing a final histogram that resembles a large Gaussian (the 0-electron cells convolved with a Gaussian) with an exponential tail (the result of CIC amplified by the EM gain). We did not add in the scaling and offset from the readout stage. In a real histogram, we would only be able to see the total histogram, shown by the black curve. However, in our model, we can color-code the results to show which parts of the histogram come from cells with certain initial numbers of electrons. We can clearly see that the main peak is mostly from the cells with no initial electrons, while the exponential tail is almost completely from the cells with only one. Contributions from higher electron counts are negligible.

The EM leakage changes the histograms so that the tails are no longer as close to an exponential, but, for small leakage, the biggest effect is in the transition from the Gaussian peak to the exponential tail. In the absence of EM leakage, this elbow is pretty sharp. In the presence of EM leakage, though, the charge produced in the EM stage gets amplified just like any other charge. Some of the charge is added towards the beginning of the EM stage, and so gets the full EM gain amplification added to it, which just adds to the apparent CIC tail. Some of the charge, though, is added towards the end of the EM stage, and gets almost no amplification at all. This charge adds a small steep tail to the histogram. The net effect of having EM leakage is we start with a fairly sharp elbow, but, as we add charge that is created further and further down the EM stage, we add steeper and steeper tails to the side of the Gaussian peak. While this actually effects the entire tail, the most obvious visual effect in the histograms is that is makes the elbow rounder. We will discuss this further in Section V.5.

The above approximation of the histograms is a useful way to picture these histograms: They are a Gaussian peak with a roughly exponential tail. Approximately, the width of the Gaussian peak gives the readout noise, the length of the exponential tail gives the EM gain, and the area under the exponential tail gives the CIC noise. Since the readout noise and EM gain are both scaled by some unknown readout gain factor, we can scale the histograms so that the readout noise is unity, which changes the decay rate of the exponential tail to the readout noise divided by the EM gain (and the readout scaling cancels out). Plotting multiple histograms this way allows us to compare the ratio of EM gain to readout noise, by looking at which histogram has the longest tail. Alternatively, if we know that two histograms should have identical readout stage scalings (such as, for example, histograms from the same camera in the same readout mode), we can directly compare the widths of the Gaussian peaks and the decay rates of the exponential tails, and look for relative changes in them.

Figure 5.6 shows measured histograms, taken with some demo cameras we evaluated, including a Hamamatsu ImagEM C9100–13 and a Princeton Instruments

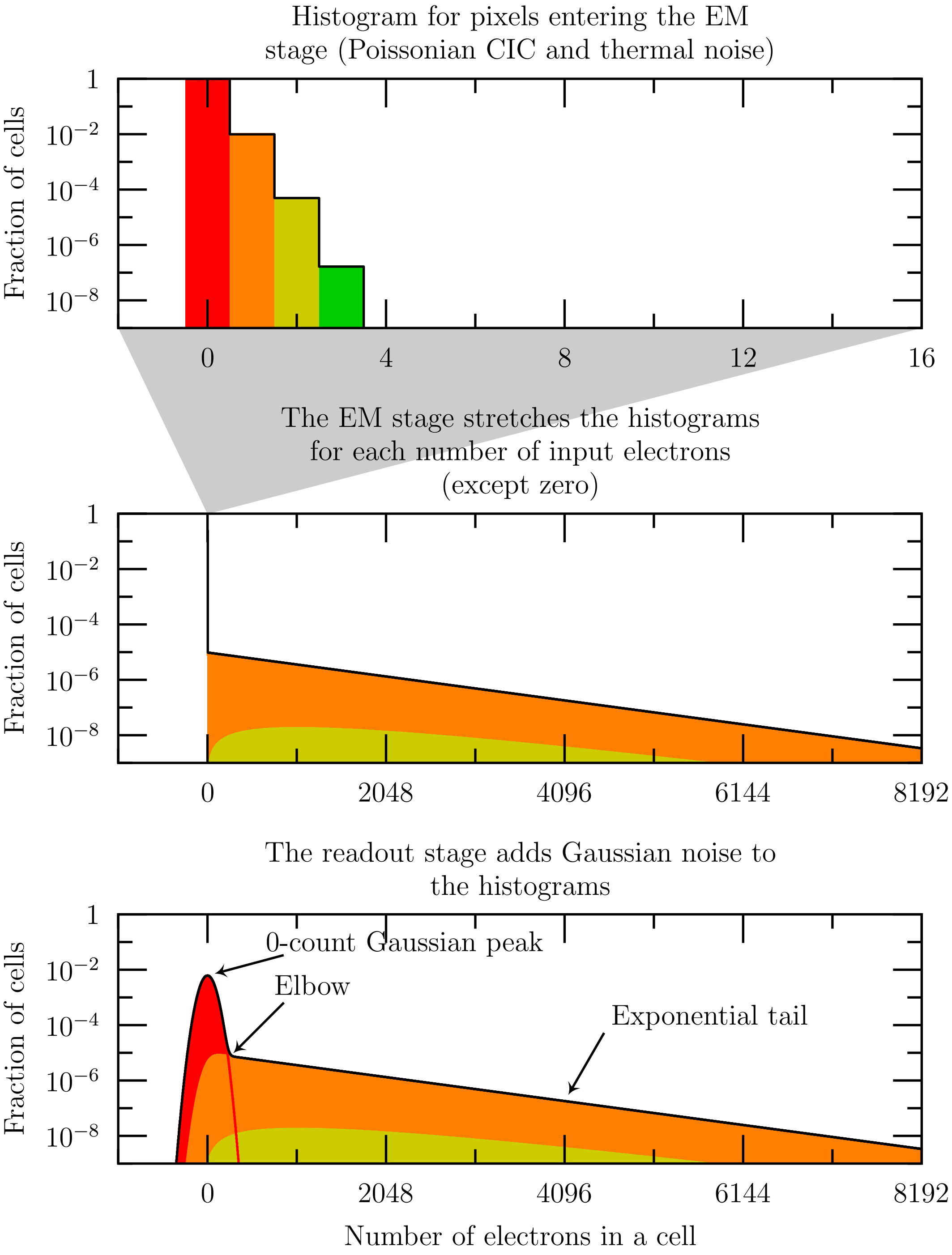


Figure 5.5. The progression of a camera histogram as it passes through our model of an EMCCD camera, approximated by Equation (A.56).

ProEM:512BK_eXcelon. The features are hard to discern, as the exponential tail is so small compared to the main Gaussian peak, so we will present histograms on a semi-log scale. Figure 5.7 shows the same histograms on a semi-log scale, where the features are more obvious. We will use semi-log scales almost exclusively when dealing with histograms.

To lend something slightly more physical to these histograms, we show two actual dark frames take from a Hamamatsu ImagEM camera. Figure 5.8 shows a dark frame with no EM gain, and Figure 5.9 shows a dark frame with maximal EM gain. Both images have a Gaussian-noise background that represents the readout noise. The EM image has some bright pixels that swamp that background noise. These pixels make up the exponential tail in the histograms from the amplified electrons. Images such as these can make one think that the amplified electrons are completely distinguished from the Gaussian readout noise, and this is close, but there is actually no fixed cutoff that separates the two, as seen in the histograms.

Before we invest too much time in discussing ways of using this model, we should first verify that it actually works. While we have not explicitly discussed it, this same model also gives us error bars on the histograms. The exact histogram gives us the fraction of pixels that we expect to have a certain value in the final image. Since we actually have a limited number of pixels, it is unlikely that a real histogram will exactly follow the model. If we assume, though, that each of N pixels has a probability distribution given by the model histogram, then the actual number of pixels with a given output is the sum of N independent, random variables that are 1 a certain fraction of the time (given by the histogram), and 0 the rest of the time. This describes a binomial distribution. As discussed in Section B.5, since the mean value is fN, where f is the fraction given by the histogram, the variance is approximately fN (since fN is usually small). If fN is enough larger than 1, then the binomial distribution is, to a decent degree, Gaussian. Thus, if we add error bars to our histogram (scaled to the full number of pixels N instead of unity) with a length of two standard deviations, for the region where the binomial distribution is enough like a Gaussian, that gives us a decent 95% confidence interval. Figure 5.10 and Figure 5.11 show some sample fits of this

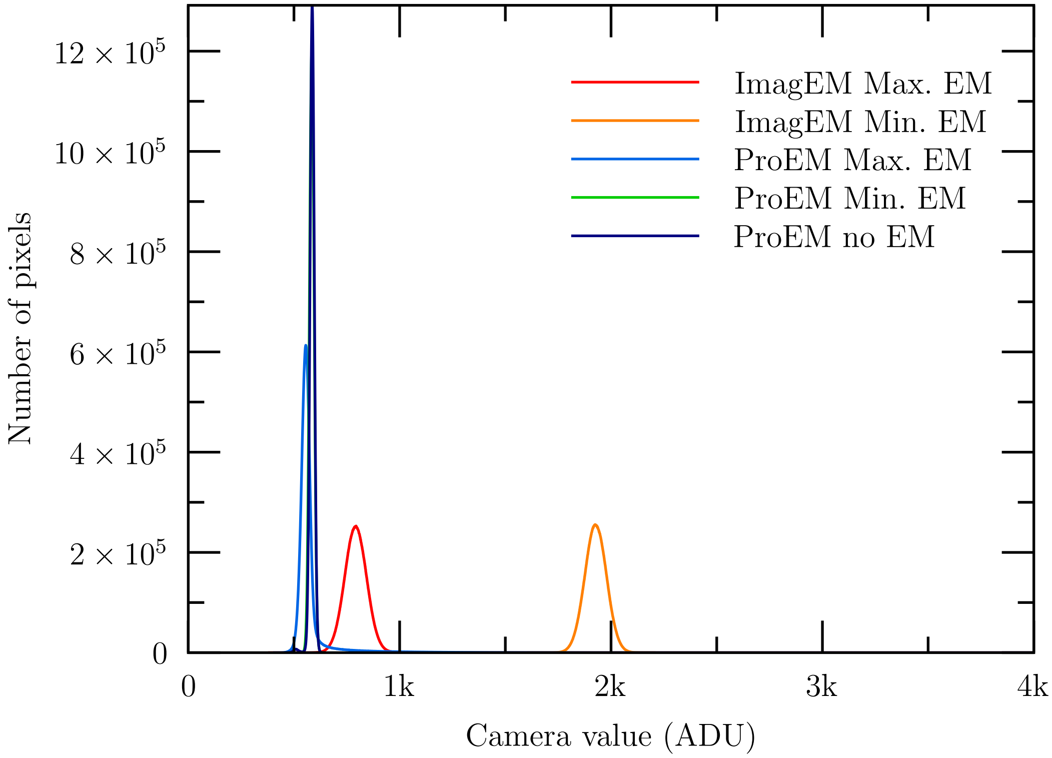


Figure 5.6. Some sample histograms from different camera models in a few different imaging modes. Each histogram is computed from 128 consecutive images taken with a single camera in a single imaging mode. We can clearly see the Gaussian peaks, but in only one histogram is the tail we expect to see from an EMCCD evident. Note that two histograms overlap each other almost completely in this figure.

model to two actual histograms, one from a Hamamatsu ImagEM camera, and another from a Princeton Instruments ProEM camera. Both fits are good; the ImagEM fit is impressive. Even the estimated error bounds hit the mark quite closely. Our model tended to produce fits of the quality shown in Figure 5.10 for the ImagEM cameras we looked at, but typcial fits to ProEM data were usually not as closely fit as the example shown in Figure 5.11. The histograms show a somewhat non-Gaussian tail, which is where Gaussian approximations from the central limit theorem are most likely to deviate. The largest problem is probably that the ProEM readout noise is not very Gaussian, and so the model could probably be improved

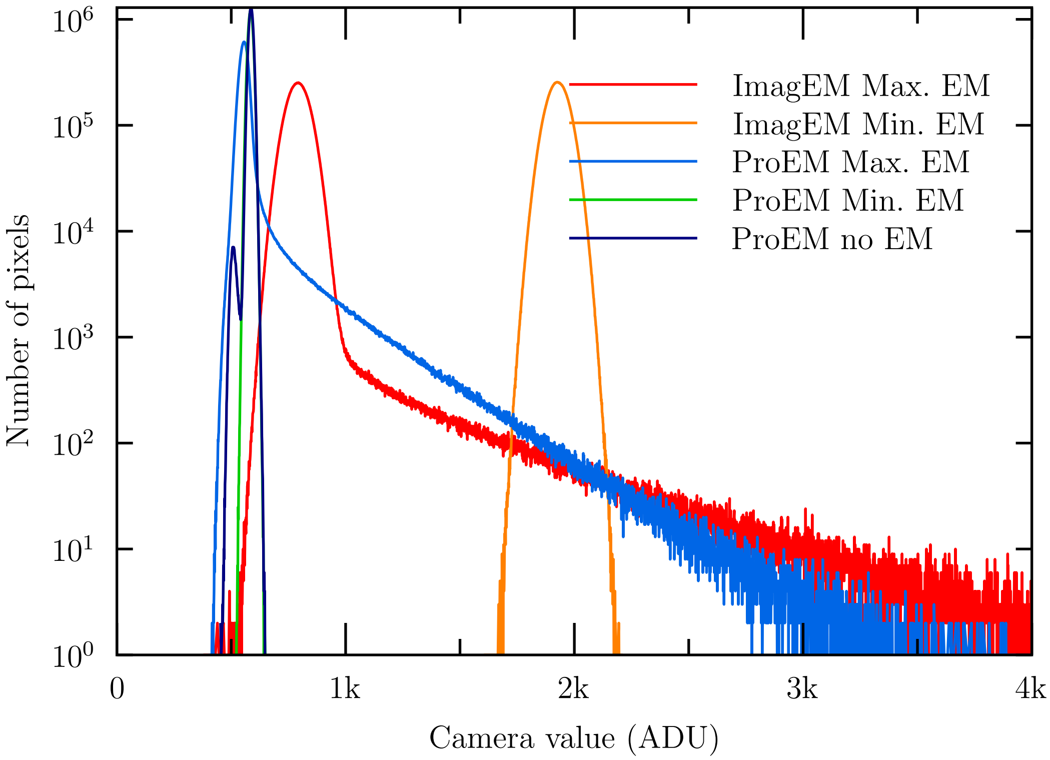


Figure 5.7. Some sample histograms from different camera models in a few different imaging modes on a semi-log scale. This shows the same sample histograms as Figure 5.6, except on a semi-log scale, which allows us to see the important features even though they span several orders of magnitude. In particular, we can clearly see how all the histograms have an upside-down parabola, (the Gaussian peak), and how the histograms with large EM gains have long exponential tails from the CIC noise multiplied by the EM gain. We can also make out a very non-Gaussian feature on the left of the “ProEM no EM” curve, which we will discuss briefly in Figure 5.14 and the referring text, and more thoroughly in Section V.8 and Section V.11.

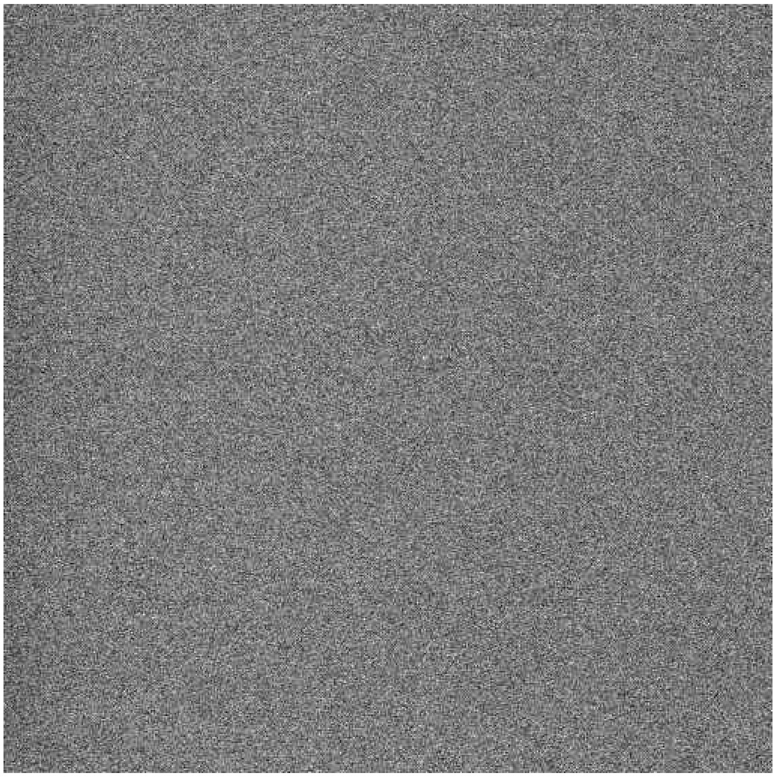


Figure 5.8. A sample dark frame from an ImagEM, with the EM gain turned off. It is dominated by the readout noise. The grayscale has been adjusted to the scale of the fluctuations, which are really quite small compared to the dynamic range of the camera, or even a faint light source.

by convolving with some not-quite-Gaussian function that more closely matches the ProEM readout noise. Determining such a function, though, could be problematic, since the actual function most likely changes with the imaging parameters. We know, for example, that the standard deviation seems to increase with EM gain, as shown in Section V.8, so it seems unlikely the function will be easily modeled.

Now we see that our EMCCD model, which was based off of mostly justified assumptions on how EMCCDs work, predicts very well what an EMCCD histogram should look like. In other sections, we will attempt to justify some of our other assumptions further, such as the existence of the EM leakage term we added, which we discuss in Section V.5. We will also make many quantitative measurements and

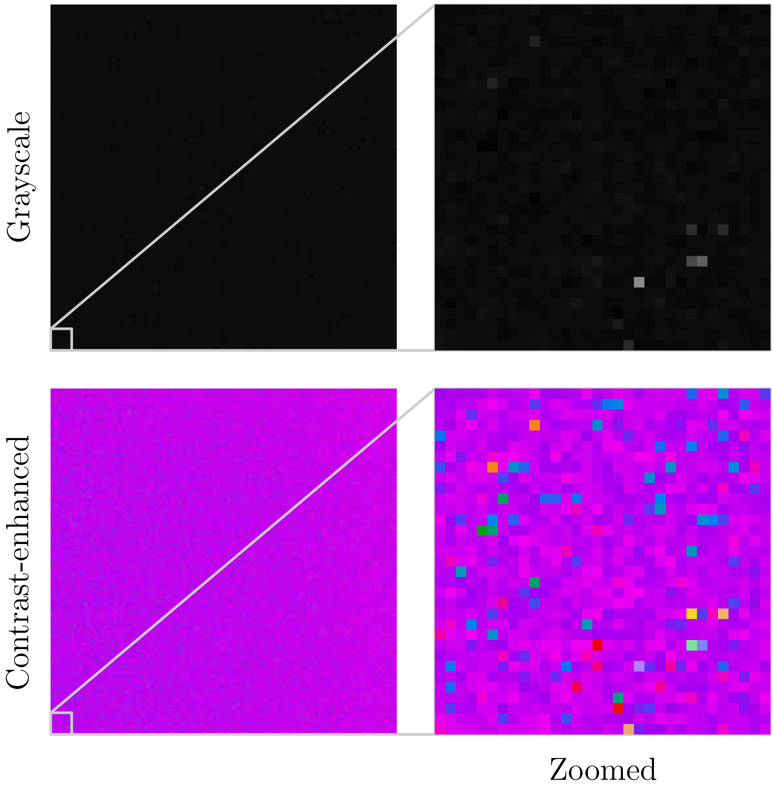


Figure 5.9. A sample dark frame from an ImagEM, with the EM gain turned up to the maximum value of 1200. The images on the left show the entire dark frame, while the images on the right zoom in on the lower-left corner. The upper images show the dark frame in grayscale, while the lower images show the dark frame using a high-contrast color scheme.

comparisons of cameras in later sections, and, where there are discrepancies with quoted values, we will attempt to explain the differences. For now, however, we will take some time to use the intuition the model has given us to compare histograms qualitatively.

One thing we learned from looking at histograms is they sometimes make certain camera problems quite obvious. We have already seen one mentioned briefly in Figure 5.7. One set of dark frames that we took with a Hamamatsu ImagEM seemed to be the best yet; a sample image from that set is shown in Figure 5.12. In these images, the background seemed nearly noiseless, and the occasional spike

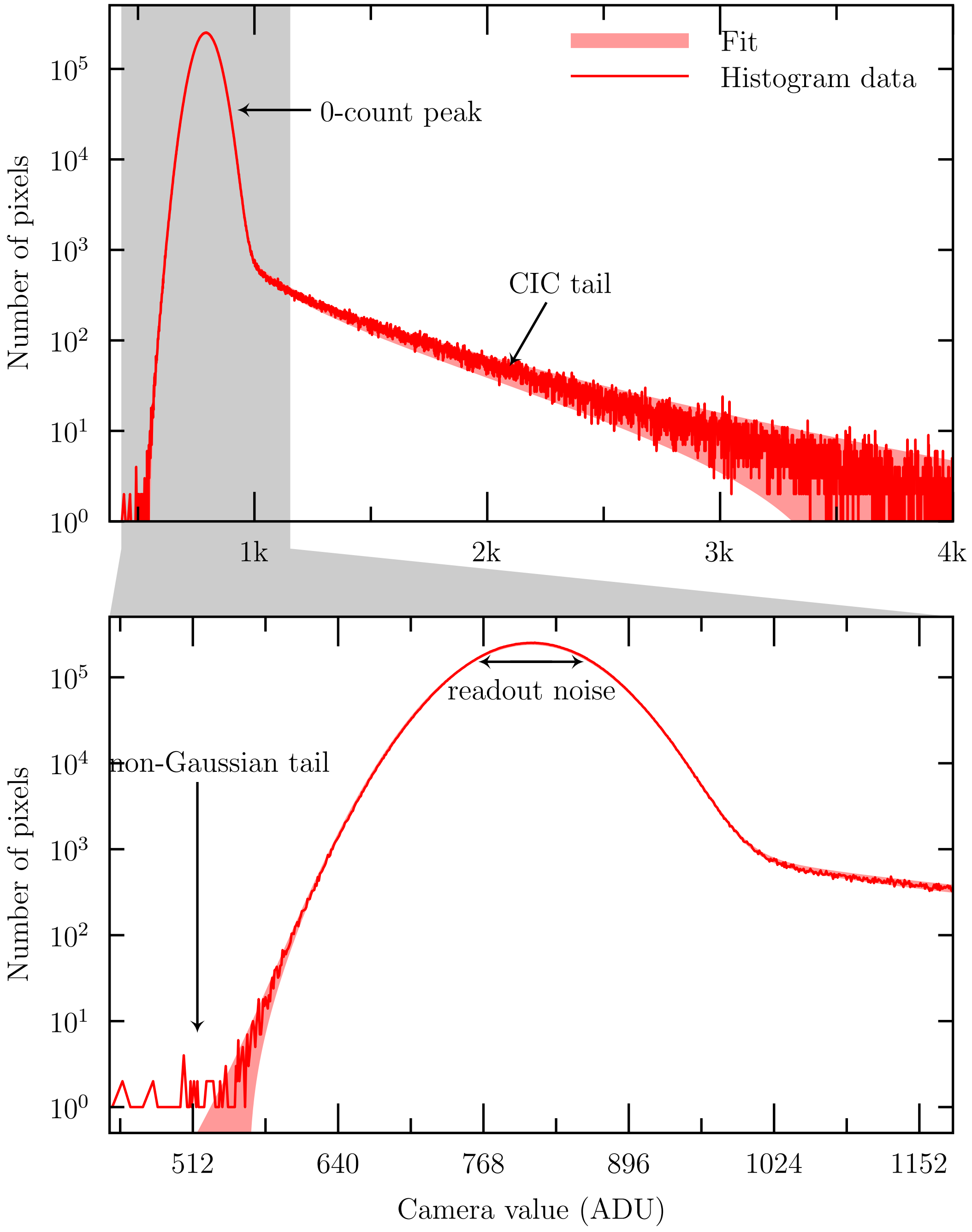


Figure 5.10. A comparison of our EMCCD noise model with an example histogram taken from a Hamamatsu ImagEM EMCCD camera. The thickness of the fit at each point represents an approximate 95% confidence interval that we expect the actual histogram to fall within.

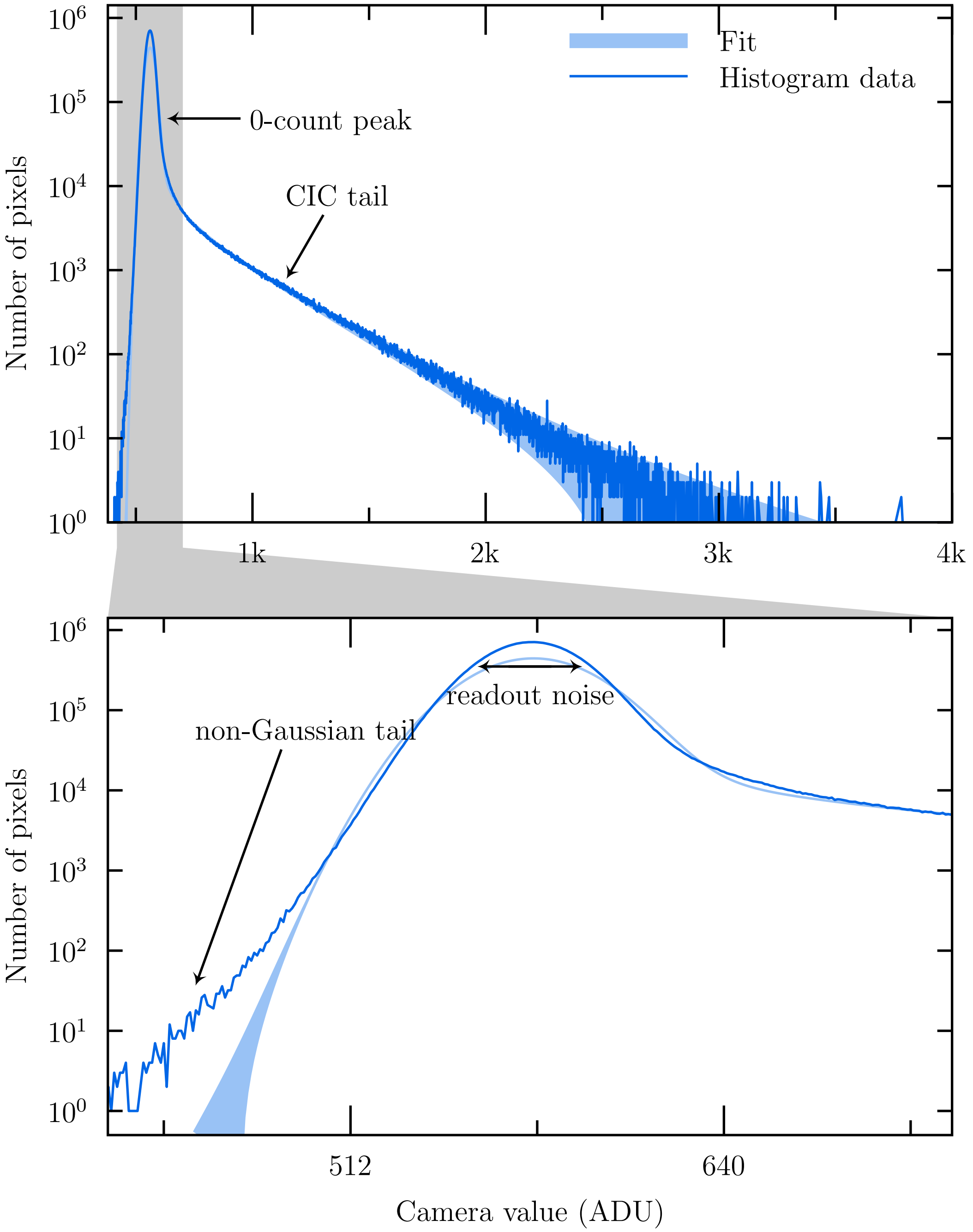


Figure 5.11. A comparison of our EMCCD noise model with an example histogram taken from a Princeton Instruments ProEM EMCCD camera. The thickness of the fit at each point represents an approximate 95% confidence interval that we expect the actual histogram to fall within.

from the CIC seemed to be amplified well above the background. Back when we saw these images, we still thought that an EMCCD would amplify a single electron into a Gaussian peak well separated from the zero-electron Gaussian peak, and these frames seemed to agree with that. In general, the histograms show that this is not the case, as the CIC tail very clearly extends from within the readout noise to well outside that range. However, we found some images that really seemed to confirm our original thoughts. The histogram for one set of those is shown in Figure 5.13, where we can clearly see that this is the result of a mistake, rather than amazing performance. The CIC tail really stands out because the background "readout noise" is almost exactly zero, but only because it is clipped off by an incorrect readout stage offset value. Not all modes of the ImagEM seem to have this problem. Normally, all the readout stage offsets seem to be fixed at 2000 ADU. However, as the EM gain is increased, this offset seems to shift, with much greater shifts at higher readout-stage gains than lower readout-stage gains. In the highest readout-stage gain, the shift at full EM is so large that almost all of the main Gaussian peak is clipped, as shown in Figure 5.13.

Thus, we see that just having a basic idea of what the histograms look like allows one to catch certain problems with certain cameras or camera modes. We now look more at comparing cameras based on this general idea of what the histograms look like.

One way to compare histograms is to plot them on a common scale to try to compare them. For instance, we can try to plot the histograms with the horizontal scale being number of electrons instead of ADU (which is different from camera to camera and between different readout modes), and shift them so that the zero-electron peak is centered at zero-electrons. Each camera from Princeton Instruments that we tried as a demo camera came with a spec sheet that gave the electrons per ADU conversion factor for each readout mode of the camera, which allows us to convert ADU to electrons for ProEM histograms. The ImagEM cameras from Hamamatsu did not come with such a conversion, but we can infer the conversion if we assume the gain is accurate. The ImagEM cameras claim a maximum EM gain of 1200x; if we assume that is accurate, then we can fit our histogram model to the

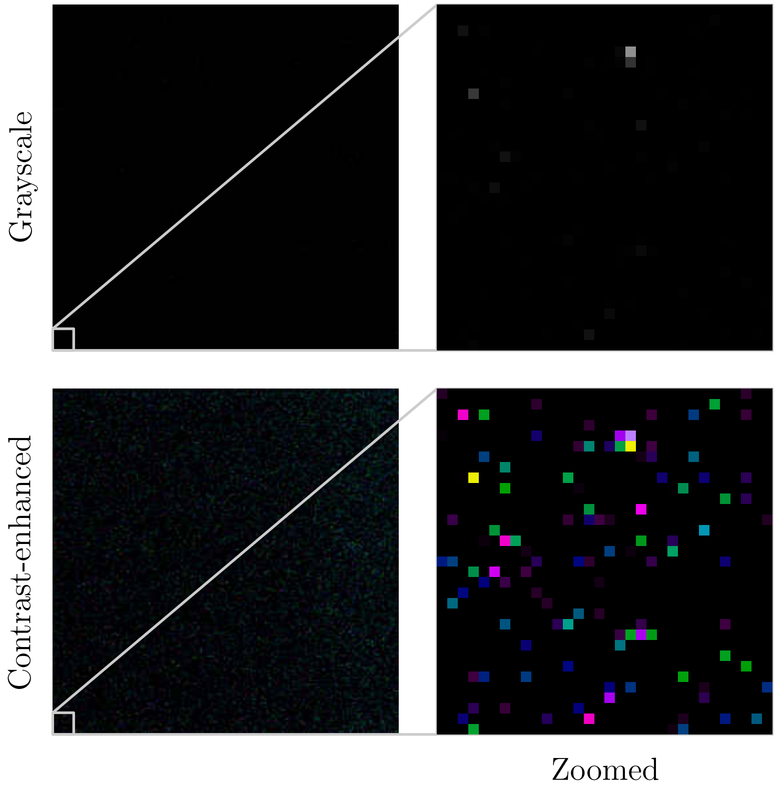


Figure 5.12. A sample dark frame from an ImagEM in a mode where the histogram gets clipped (as shown in Figure 5.13). The images on the left show the entire dark frame, while the images on the right zoom in on the lower-left corner. The upper images show the dark frame in grayscale, while the lower images show the dark frame using a high-contrast color scheme. Note how the images appear to have a completely flat background with occasional spikes.

histogram, which gives us the EM gain divided by the readout stage scaling, which we can divide into the actual EM gain to get the electron to ADU conversion. If we further assume that the electron to ADU conversion is independent of the EM gain, we can use the same scaling to plot histograms with low EM gain, but not non-EM-gain modes, as those have different readout electronics and probably different scalings. This same method for inferring the readout stage scaling on the ProEM cameras often agreed with the given specification to within 10% percent (although there were discrepancies), and it typically gives readout noises near the quoted

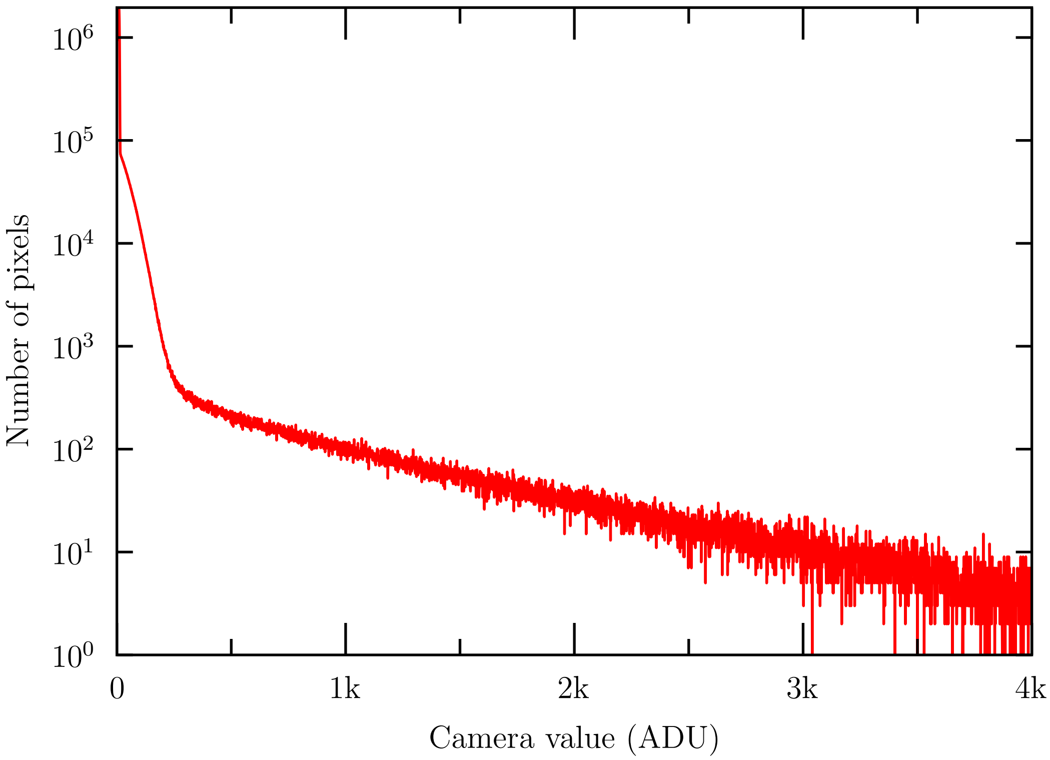


Figure 5.13. A clipped dark-frame histogram from an ImagEM. The readout stage applies a DC offset to the readout to shift the zero-electron Gaussian so that the entire Gaussian can be seen. In this histogram, from a series of dark frames from a certain mode of a Hamamatsu ImagEM, we see that the offset was not correctly set. Here, we can clearly see that the Gaussian has been clipped, and a large number of pixels have been put in the 0 ADU bin. The dark frames associated with this histogram, shown in Figure 5.12, actually look great, as the CIC spikes seem very clearly differentiated from a seemingly noise-less background. One might think that the random spike was just amplified to well above the background noise, but here we see the background is seemingly noiseless for a much more problematic reason.

typical values, so we will treat this value as tentatively accurate, but not base any important decisions on it. Figure 5.14 shows more histograms plotted with this new scaling. We immediately see that the readout noise of the two ImagEM imaging modes does not change as the EM is increased from its minimum value (where we cannot see any amplified tail) to its maximum value. The ProEM, however, shows large changes in the readout noise. The first change is when switching from a no-EM mode to an EM-mode. The camera specifications state that the typical readout noises for these two modes are different, and so we expect that [102]. There is a large change in readout noise between the low-EM and high-EM histograms, however. We suspect that this is due to a coupling between the readout stage and the EM stage, and appears to be mostly an actual increase in readout noise, and partly an increase in readout stage gain (which we do not account for here as we are using the value given in the specification sheet). The ProEM has less readout noise than the ImagEM, and so we would expect other noise sources to become more apparent, but there is no increase of this magnitude in the ImagEM histogram. We also see some large bumps to the left of some of the zero-count Gaussians in the ProEM histograms. These appear to be due to an odd bug where the first image of the 128 image sequence tends to have a different readout offset (and even a changing readout offset) than the rest of the images; this bug is discussed in more detail in Section V.8 and Section V.11.

As mentioned, we can use a different scaling of these histograms to aid in comparing them. Figure 5.15 shows two high EM gain histograms from two different cameras, horizontally scaled so that the readout noise is unity. The vertical scaling was adjusted to preserve the area under the two curves, which are both $512 \times 512 \times 128$ pixels, since these were made from 128 images of 512×512 pixels each. With such a scale, we the slope of the tail tells us the EM gain divided by the readout noise, which is a measure of how easily we can distinguish the tail from the main Gaussian. The fraction of area under the tail tells us what fraction of the pixels had CIC electrons in them, which is a measure of the CIC. The first ProEM image of the sequence is included in the histogram, and was a little off (see Section V.11 for more on this), so we did not use it when computing the scalings to apply

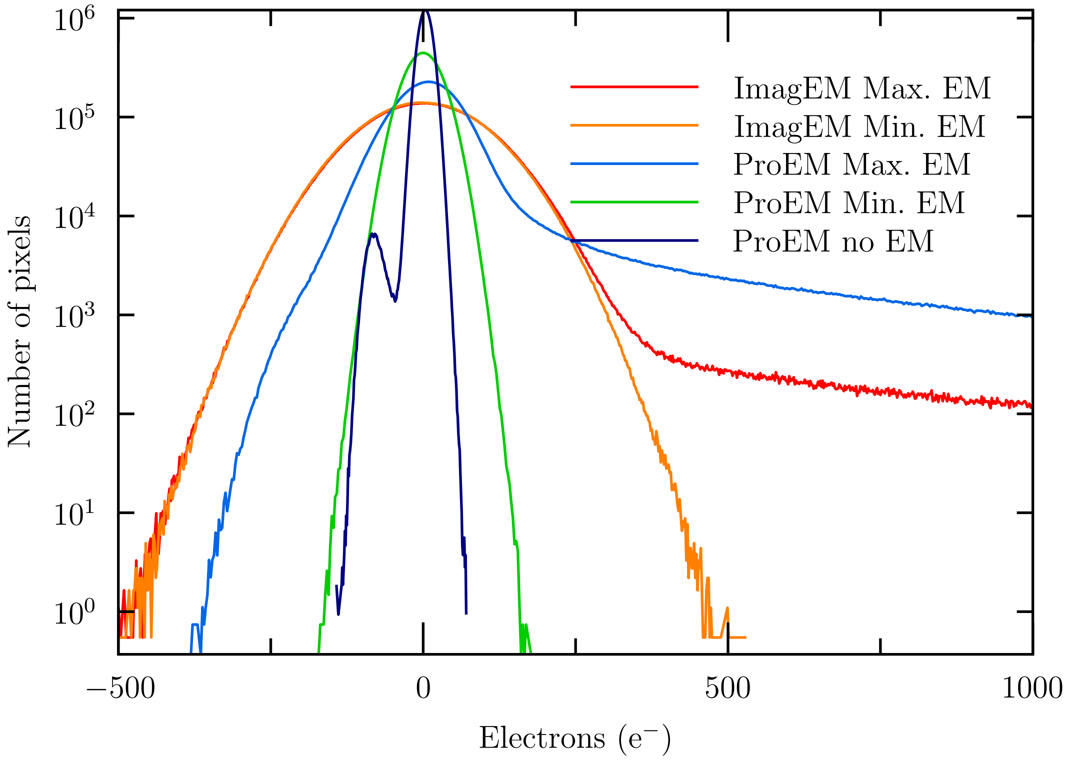


Figure 5.14. A comparison of sample dark-frame histograms. These are the same sample histograms used in Figure 5.6, but scaled back from ADU to electrons, and shifted so that the peak of the Gaussian corresponds to zero electrons. The readout stage scalings are inferred as described in the text. Here, we note that the ImagEM camera readout noise appears not to change as we increase the EM gain from the minimum value to the maximum value, as seen by how the width of the Gaussian does not change (and just a EM-amplified CIC tail appears). The ProEM histograms show other effects. First, two of the histograms have a side bump to the left of the main one. These result from the first-in-sequence problem described in Section V.8, where the first image of the sequence has a shifting offset or other problem. We describe this problem in greater detail in Section V.8 and Section V.11.

to these histograms.

We can see a slight difference in the slopes of the tails in Figure 5.15. The ProEM has a longer tail than the ImagEM, which implies it has a higher EM gain after dividing by the readout noise. The ProEM actually has less EM gain overall than the ImagEM, but the readout noise in the ProEM, even with the increase as EM gain increases, is more than enough to compensate. Fitting the histograms confirms this, but, as seen here, the difference is slight. We do, however, see a huge difference in the relative tail areas. Ordinarily, this is a little difficult to compare, but when the slopes of the tails are about the same as they are here, we can judge it pretty well by eye. The ImagEM has approximately an order of magnitude less CIC than the ProEM, and less EM leakage, which we can tell because the elbow is much rounder on the ProEM histogram.

Fitting this histogram model to actual histograms gives us fairly reliable noise information about the cameras, and allows us to compare the cameras in an effective manner. We will mention our results from doing this throughout the rest of this chapter.

Defending the Existence of Leakage in the EM Stage

We have shown that our model for an EMCCD histogram fits real histogram remarkably well, and that the largest problem seems to be that the Princeton Instruments readout amplifier produces non-Gaussian noise, but that the Gaussian approximation is still decent. We have also demonstrated that the model allows us to compute and compare camera specifications remarkably well, which we will do in much greater depth later in this chapter. We can even, from intuition gathered from the model, estimate and compare specifications of the cameras by eye, and even detect some anomalies, all by simple inspection of the histogram of an EMCCD camera. While all of this suggests that our model of an EMCCD camera is a good one, there remains one odd issue: EM stage leakage.

In discussing the model, we simply postulated that extra charge was produced during the EM stage. The process and statistics were assumed to be just like CIC.

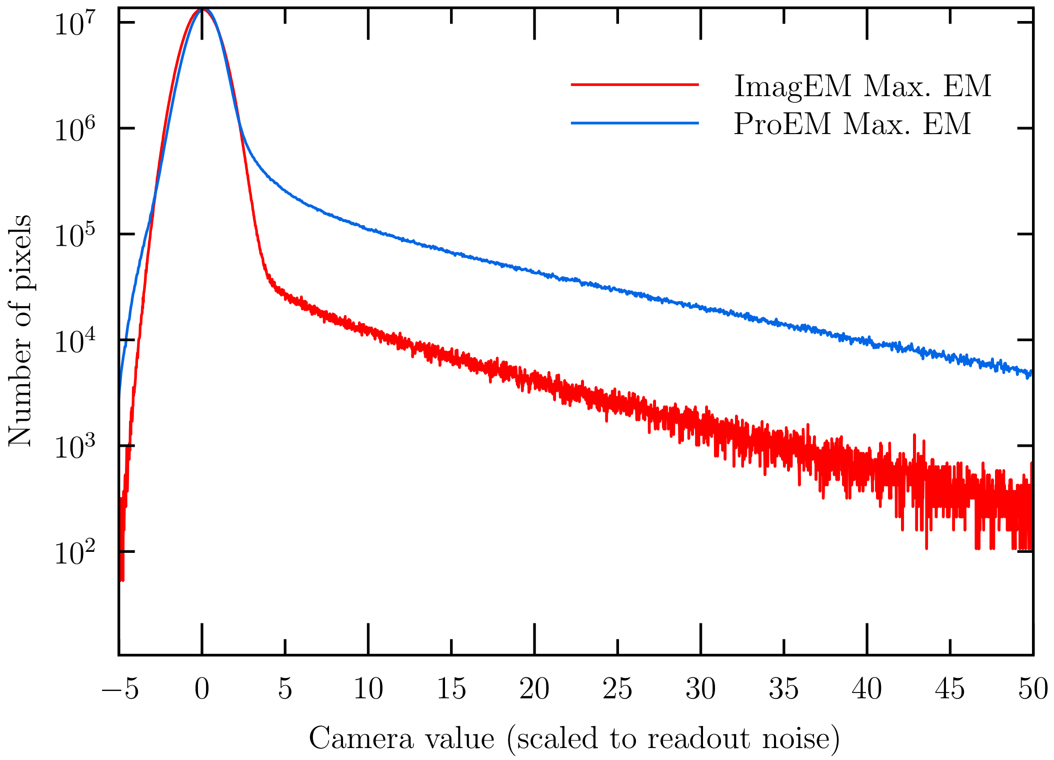


Figure 5.15. A comparison of dark-frame histogram tails. These are the two maximum EM gain histograms shown in Figure 5.6, but scaled so that the Gaussian peaks have the same width (and vertically scaled to preserve areas unde the histograms). The same oddity on the left side of the Gaussian peak of the ProEM histogram discussed in Figure 5.14 can be seen here, but the image that creates that bump was removed when computing the scaling to apply to the histogram. In this view, we can easily compare the CIC noise and EM gain to readout noise ratios. The EM gain to readout noise ratio, when the readout noise is scaled to unity, determines the basic slope of the tail on a semi-log plot. Here, we can see that the ProEM has a longer tail (less steep), implying a larger EM gain when scaled down by the readout noise. The CIC noise is represented by the area under the tail (as a fraction of the total area under the curve, which is the same for the two histograms). The ImagEM has about an order of magnitude less CIC, and less EM leakage, as seen by how much sharper the elbow between the Gaussian peak and the exponential tail is on the ImagEM histogram than the ProEM histogram.

CIC for the horizontal shifts is typically less than for the vertical shifts, and we were told that this held in the EM stage as well [40, 41]. Nevertheless, we chose to include a term like that for the EM stage in our model. We reconcile this with the no-horizontal-CIC statement by assuming that it is not so much a clock-induced-charge process as an amplification leakage process. There is a finite energy gap between the valence and conduction bands of the semiconductor substrate of the EMCCD, and so, since the temperature of the system is non-zero, there is a non-zero probability that some charge will spontaneously be elevated to the conduction band. Since the EM stage is a gain stage with larger voltages applied across it, and these voltages accelerate electrons across the cell with enough energy to promote other electrons to the conduction band, it seems reasonable to assume that this spontaneous elevation would be greatly enhanced. In fact, this is very similar to the spontaneous "dark counts" generated by standard photomultipliers, and seem to be an amplification of the thermal noise that is regarded as a problem during the exposure itself. However, we are still adding something to our model that does not seem to be an accepted effect in the EMCCD community, so we feel some extra justification of this effect is needed.

Our first attempt at a model actually did not include an EM leakage term. Figure 5.16 and Figure 5.17 show the same two histograms as Figure 5.10 and Figure 5.11, respectively, with fits of the same model to those histograms. The difference is that the fits shown were forced to have no EM leakage, and show the sort of fits we were getting with our first attempt at an EMCCD model. Note that the fits reproduce the general form of the histograms, but really miss parts of them, with nowhere near the accuracy as with the full model. In particular, the fits struggle to match the CIC tail, particularly with the ProEM data in Figure 5.17. Without EM leakage, the fits tend to produce an exponential tail with a rather sharp transition between the Gaussian peak and that tail. The actual histograms have a transition that is a little less sharp. When we do not include the EM leakage, our remaining parameters can only change the vertical offset and the slope of the exponential tail, and so cannot match the curved tail. We demonstrate this in Figure 5.16 by showing an extra fit, where we emphasized trying to match the tail, but missed the transition as a result.

For clarity, this fit is shown without the confidence interval.

We considered what effects we may have glossed over that could affect the histograms. It is possible that the mathematical approximations we made in deriving the formulas for the model were mistaken. However, for most of those approximations, we were able to show directly that the corrections were very small. As for the rest, we compared our final formulas to numerical computations that did not have those approximations, and found that they agreed with each other quite closely. The problem did not seem to be a mathematical approximation or error, which left how we were modeling the physical system, and the approximations we were making of those physical elements.

We had already included all the physical parts of an EMCCD in the model, and accounted for CIC, thermal noise, and readout noise. The large peak in the histogram pretty much shows the readout noise shape, which was not deviant enough from a Gaussian to explain the differences we saw. The CIC and thermal noise might not be Poissonian, but that seemed unlikely. Any discrete random process with such a low mean is almost guaranteed to be nearly Poissonian, simply because such processes tend to produce either 0 or 1 as a count, with 0 being much more common than 1, and that is a good approximation of a Poissonian distribution with a low mean. Also, we see no reason to expect CIC and thermal noise to have any sort of dependence on the number of electrons already in a cell, until the cell approaches saturation, but even if there was a dependence, since the rates are low enough that they should just be producing 0 or 1 electron the vast majority of the time, there are not enough 2 events that the dependence should make a difference.

We came up with a list that was basically the mostly likely effect we were missing at each stage of an EMCCD:

1. On the main CCD, something could change as a function of which cell of the CCD was being read.

2. In the EM stage, the gain could be changing.

3. In the EM stage, extra electrons could be gained or lost, regardless of presence

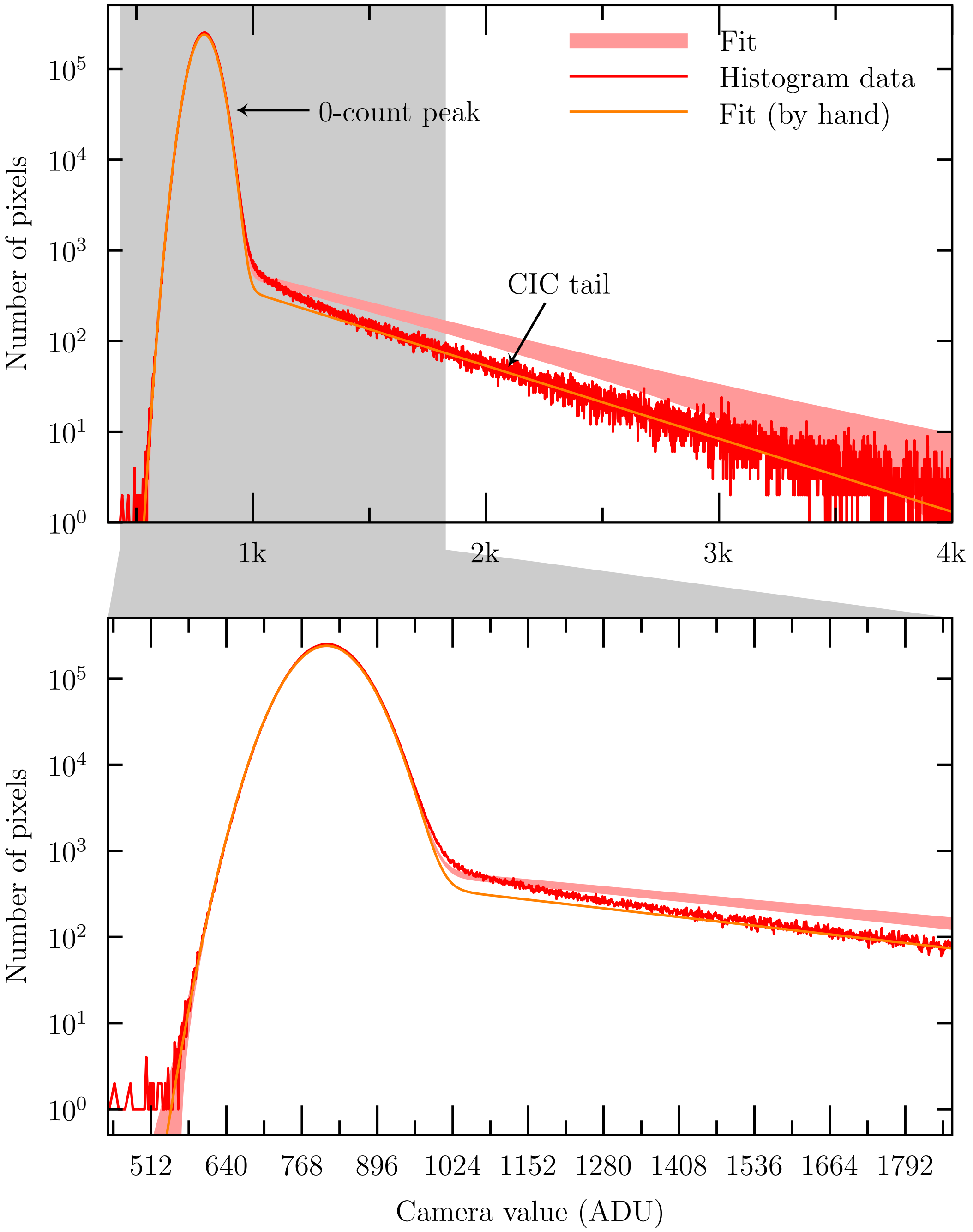


Figure 5.16. A comparison of our EMCCD noise model without EM leakage with an example histogram taken from a Hamamatsu ImagEM EMCCD camera.

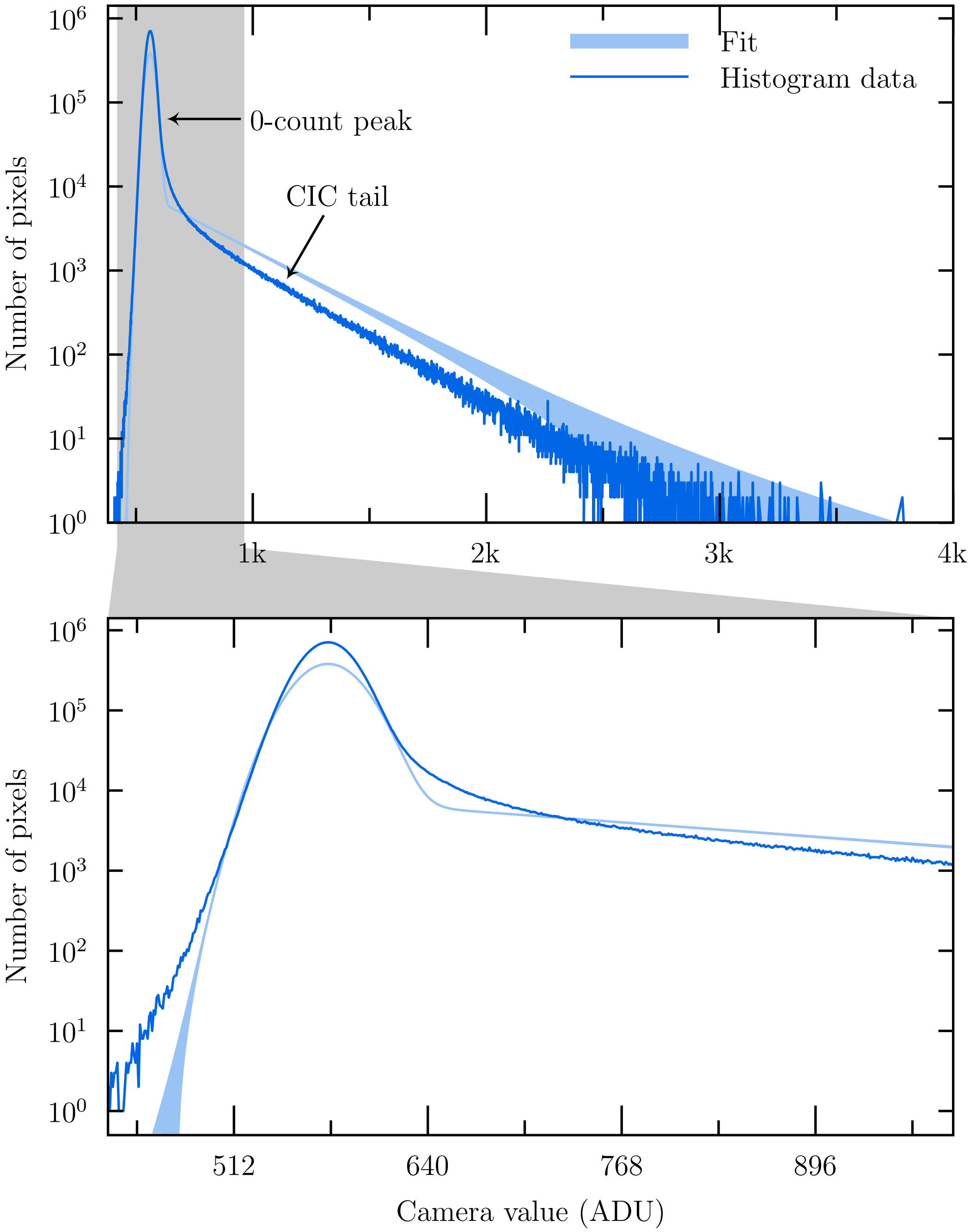


Figure 5.17. A comparison of our EMCCD noise model without EM leakage with an example histogram taken from a Princeton Instruments ProEM camera.

of electrons already there.

4. The readout stage could be nonlinear.

Let us consider what would happen to the histogram if something changed during the readout, perhaps as a function of what was being read. If the readout offset changes, that is equivalent to adding a certain number of electrons to the readout, which is what the readout noise is. Thus, this would effectively be a change in readout noise.[3] If it were the amount of CIC and thermal noise that was changing, as long as those values remained small enough that 2 or more electrons is rare, the pre-EM probability distribution would still be nearly Poissonian (since it is just mostly zero electrons with a small one-electron probability), with a mean that is just the average of the means of all the pixels. The readout stage gain could change as a function of position, but this seemed unlikely based on the ImagEM histogram. As seen in Figure 5.10, the main peak is very Gaussian. If the readout noise is very Gaussian, then adding together Gaussians of different widths would tend to flatten the sides of that peak in a non-Gaussian way. Also, to have a significant change on the peak, the fluctuations have to be fairly large (on the order of 10% or more). The peak is a decent representation of the function that we should convolve with to represent the readout noise, so if the peak is not visibly affected, then the convolving function (and hence the entire distribution) should not be greatly affected. While it is possible that the readout noise was not Gaussian, but changed in such a way that the aggregation *was* Gaussian, that seemed unlikely.

Pretty much the only parameter left to vary over the CCD is the gain, which brings us to the second point. There are two ways the EM gain could change. The net EM gain could vary from pixel to pixel (or slowly across the entire image), which could be thought as as belonging to the first point as well, or the gain of each step in the EM stage could vary. The second case is rather simple. In deriving the formula for these histograms, we argue that there are so many stages to the EM

[3] This actually does happen. Shading is basically an example of this. When it happens reliably as a function of position on the CCD, you can remove the effect by subtracting two images. If it is not so reliable, it effectively just increases the readout noise.

stage that it is very close to continuous.[4] In particular, in Equation (A.25) we show that in the continuous case, it is really only the average value of the gain in each step (which determines the overall EM gain) that affects the final result, and not the variations in the gain, especially if you are looking on a logarithmic scale as we do. Therefore, when the EM gain changes within the stage, the effect only changes the histogram through changing the total EM gain.

The EM gain could change from pixel to pixel. EM leakage is, in a sense, CIC with variable gain (depending on where in the EM stage the leakage happened), and since that fits well, we could expect a changing EM gain due to other reasons might fit the histograms we see. One difference, though, is that EM leakage results in an effective gain *only* for the charge additions that happen in the EM stage, whereas a truly varying gain would affect all of the electrons. Another important difference is that EM leakage covers the full range of EM gain, from none (the charge is produced at the end of the EM gain stage, and so does not get amplified), to full (the charge produced at the very beginning, and so traverses the entire EM gain stage), while we would expect fluctuating or drifting electronics to alter the gain in some small range about a fixed mean. Since dark frames do not have any signal, that difference would show in the CIC. Figure 5.18 compares histograms with no EM leakage and possibly some EM gain variations to an actual histogram that the EM leakage model fit well (as shown in Figure 5.10). The model fits well, but requires very large fluctuations in the EM gain. For the fit shown in the figure, we assumed the EM gain fluctuations had a standard deviation of about 40% of the mean gain value, which seems unbelievably large. We can make sense of this fit by comparing it to EM leakage.

EM leakage, as we have shown, fits the histograms quite well. As we just mentioned, EM leakage is a little like having a small amount of CIC that gets amplified by different amounts. Thus, you can imagine taking a series of no-leakage of his-

[4] We also numerically tested this for EM gains of around 1000, for around 500 steps, which is about what these cameras should have. We found that the continuous-gain model was indeed very close to the discrete version, even when the gain for each step oscillated by on the order of 10%. Variations in the gain do make a difference in the discrete case, but we are close enough to the continuous case that they are too small to concern us.

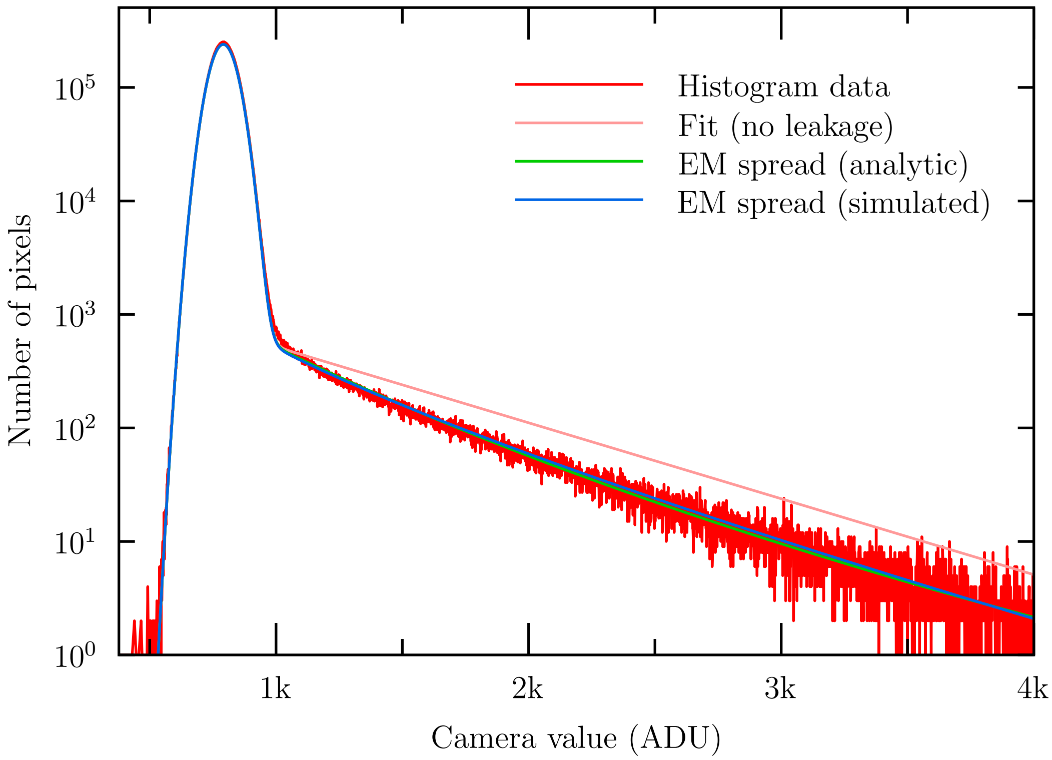


Figure 5.18. The effect of varying gain on an EM histogram. This figure shows the same data as Figure 5.16, with the same fit that assumes no EM leakage. For clarity, we have left out the confidence intervals on the fit, showing just the centerline. We have also included two other fits, where the parameters were tweaked by hand, but with a spread in EM gain as described in the text. One of the two fits uses some analytic approximations to determine what a histogram would look like with a range of EM gains, and the other numerically combines the no-EM-leakage model for a range of EM gains. The two methods of simulating a range of EM gain seem to agree with each other well, and fit the data quite nicely. The only problem is to get such a nice fit, we needed to assume the EM gain varied by about 40%.

tograms, each one looking like a Gaussian with an exponential tail, but with a range of EM gains. The low-gain ones will have a short tail that decays rapidly (steep negative slope near the elbow), while the high-gain ones will have a long tail. Adding these together results in rounding out the elbow, as seen, for instance, in Figure 5.19, and seems to be precisely the sort of rounding we are looking for.

With this description of how EM leakage rounds out the tail, we can understand how a hugely varying EM gain can fit the histogram in Figure 5.18 as well. A small variation in EM gain can bend the tail a little, but to really round out the elbow, we need some steeply sloped tails. EM leakage gets these tails by actually having some electrons experience an effectively low EM gain. A varying EM gain can get these tails by having large fluctuations, so that a small fraction of the pixels really do see a very low EM gain. Because the varying gain also has a few pixels see a much larger EM gain, the tip of the tail has a lower slope than the rest, which causes the tail to look slightly bent upwards in the logarithmic scale used in Figure 5.18. The EM leakage model has a cap on the EM gain, which results in a tail that is closer to an exponential curve, which appears straight in Figure 5.18.

We investigated one more option for the cause of the curvature, which was readout stage nonlinearity. Ordinarily, we assume that the output of the readout stage is just some gain g times the number of electrons in the pixel, plus some offset o:

$$\text{ADU} = g * \text{electrons} + o.$$

If we assume that there is some nonlinearity, then that means the gain g would change a little depending on the number of electrons. The lowest-order correction would add some small electron-number dependent correction to g, such as this:

$$\text{ADU} = g\,(1 + \epsilon * \text{electrons}) * \text{electrons} + o.$$

Figure 5.20 compares a fit of this equation to the same fit to a real histograms used in Figure 5.16. The fit is almost as good as our full fit in Figure 5.10, which we have not shown as they would nearly overlay each other. To get this fit, we used ϵ on the order of 0.02%. Having $\epsilon > 0$ means the gain is larger for larger electron values, which tends to the stretch the downward-sloping CIC tail to the right, stretching

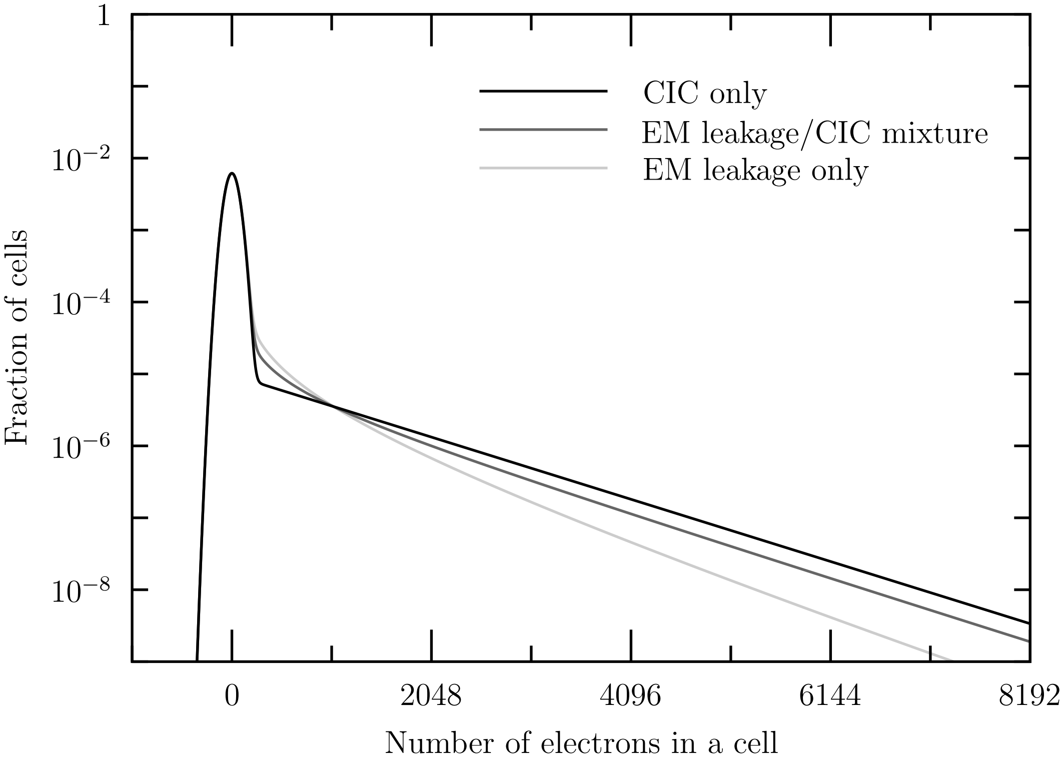


Figure 5.19. The effect of EM leakage on histograms. This plot contains three theoretical histograms with the same average CIC value, but with that value distributed between differently between pure CIC and EM stage leakage, which could be thought of as CIC that happens during the EM stage. In one histogram, the CIC tail is entirely from CIC. In another, the tail is entirely from EM leakage, and in the third, there is a mix of the two. Comparing these, we see how the EM leakage tends to round out the elbow between the main Gaussian peak and the exponential CIC tail. Harder to see is that EM leakage also make the tail less exponential; in the semi-log scale above, they have a slight curve to them.

the elbow out a little. Even though the fit is almost as good as the fit assuming EM leakage, we do not feel readout nonlinearity is a decent explanation. Initially, we thought we were on to something, since the specifications listed the nonlinearity as being less than a percent or so, which agreed with our ϵ value. We soon realized this was unlikely to be correct. Our fit has a nonlinearity of about 0.02% *per electron*, instead of the range of the detector. The measured value, in ADU, for $10,000$ electrons would be (after subtracting the constant offset) about three times what it would be for $5,000$ electrons, instead the double value we would expect from a linear detector. That is about a 50% error over a medium range for the detector (which can handle on the order of $100,000$ electrons), which we believe is a strong violation of the quoted nonlinearity of at most a percent [101, 102]. It is possible that there are higher order corrections to the nonlinearity that make it more linear for higher numbers of electrons, but it seems unlikely to us that a readout amplifier would be *more* nonlinear at smaller values. Nonlinearity usually kicks in at high values, where electronics are more nearly saturated. Also, while the nonlinearity rounds out the elbow, the same effect bends the tail so that it curves slightly upwards. As you can see in Figure 5.20, this curve is noticeably different from the actual histogram tail, and it is unlikely that the nonlinearity stops right after the elbow. Furthermore, in order to get the nonlinear fit in Figure 5.20, we had to decrease the apparent EM gain value by almost a factor two. With the EM leakage fit, we can pick a reasonable analog stage gain to convert ADU back to electrons, and find our readout noise and EM gain very closely match the specified values for the ImagEM camera we were using. With the nonlinear readout stage fit, our EM gain is off by almost a factor of two, which seems unlikely. Thus, while readout nonlinearity provides a decent fit to the dark-frame data, it requires a huge nonlinearity that contradicts how well CCDs can measure larger signals and causes us to measure values that are much worse than the specified values.

To summarize, we developed a model of how EMCCDs work using just the standard descriptions of the EMCCD processes. Fits using these models, shown in Figure 5.16 and Figure 5.17, were unsatisfactory. We thought of several corrections to our model, both in our mathematical approximations and physical approximations. We could only think of one correction that could provide a decent fit to real histograms without assuming something obviously unphysical or somehow contradictory to other results. That single correction produced impressive agreement with the given histograms, and continued to give results consistent with our expectations even used outside of simple dark frames (like the measurement of quantum efficiency discussed in Section V.10, where the CIC measurement from the fits performed as we expected it to). Not only that, but the addition to our model was simply an

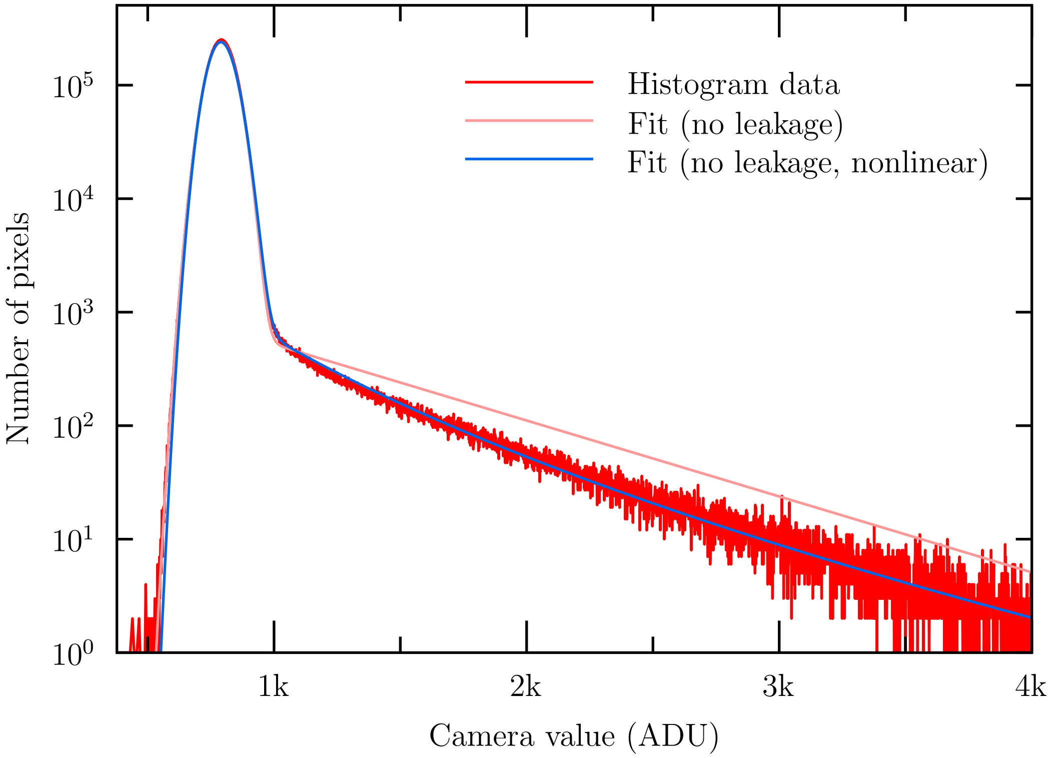


Figure 5.20. The effect of readout nonlinearity on histograms. This figure shows the same data as Figure 5.16, with the same fit that assumes no EM leakage. For clarity, we have left out the confidence intervals on the fit, showing just the centerline. We have also included a second fit, where the parameters were tweaked by hand, but with a readout nonlinearity term as described in the text. The fit is surprisingly good, but requires a very large readout nonlinearity to fit well.

effect that is common in other photo-measurement devices with near-photon sensitivity. Thus, we suspect that our addition to the model is reasonable and probably a real effect; however, we allow that our guess as to the cause (enhanced thermal excitation as opposed to horizontal CIC) may be incorrect.[5]

[5] As an aside, the thermal excitation idea could be tested by trying to measure the EM leakage very carefully over a small range of temperatures, to see if it is more sensitive to temperature than the other parameters of the model. The effect is probably also dependent on the amount of EM gain, so a similar test varying the EM gain could also be instructive. The effect probably disappears entirely at no gain, so an attempt to measure the leakage in the absence of gain, which we use as a parameter, would almost certainly not agree with the values that give good fits at high EM gain. One could also try to determine if the effect was real by somehow eliminating CIC (or fitting CIC versus number of rows, and finding the zero-row intercept), perhaps by somehow

Comparing EMCCD Cameras Using Specifications

Comparing the EMCCD cameras by their specifications is difficult. There are not many manufacturers of EMCCD cameras that we could find at this writing, and so there were not that many to choose from. Restricting the search to back-illuminated back-thinned EMCCDs with high quantum efficiency at 780 nm (the wavelength of light we are interested in) reduced it even further, and we eventually ended up with five likely cameras:

1. The Hamamatsu ImagEM C9100–13 (ImagEM)
2. The Princeton Instruments ProEM:512B_eXcelon (ProEM)
3. The Andor iXon$_3$ DU–897–CS0 #BV (iXon)
4. The Photometrics Evolve:512 (Evolve:512)
5. The QImaging Rolera Thunder (Rolera).

Even though all the cameras have the same basic EMCCD chip inside (made by E2V), the specification sheets made the cameras sound quite different [40, 101, 102, 105]. The chip itself has an exposed area consisting of a 512-by-512 array of square pixels 16 μm on a side. It is a frame-transfer chip, so there are a total of 1024 rows, but only 512 of them are exposed. The quantum efficiency is always given as a graph, as quantum efficiency is highly wavelength dependent, but all of the specification sheets agree that the quantum efficiency is about 75% at 780 nm wavelength, which is the wavelength we are using (perhaps a little higher for the ProEM with the eXcelon coating).

Due to a few misunderstandings and a reluctance to try the Andor iXon due to its Andor-specific PCI controller card, and the fact that we were unaware QImaging and Photometrics made EMCCD cameras at the time of the majority of our testing, we focused almost entirely on the ImagEM and the ProEM. While we did not try

dumping all charge right at the beginning of the EM stage, or building an EMCCD with *just* an EM stage (no rows to clock over means no CIC). In such a case, any histogram tail could only be an effect of EM leakage.

any demonstration iXon, Evolve:512, or Rolera cameras ourselves, we did manage to get some dark frames for analysis from representatives for Andor, Photometrics, and QImaging, allowing us to present some measurements from these cameras, but not to delve into whether the cameras had any anomalies or to investigate the few we did find. The noise statistics from the images we analyzed suggest that the iXon and Evolve:512 may qualify as the best camera, but we do not have the experience with them to say whether they have any odd anomalies.

The ProEM communicates via an Ethernet protocol, has something called *Kinetics mode* where the camera can take a series of very fast exposures, and has a special eXcelon coating designed to reduce interference fringes when imaging in the infrared. The specification sheet lists five readout speeds, ranging from 100 kHz to 10 MHz (but only 5 MHz and 10 MHz are available in EM mode). The quoted readout noise values are very low, ranging from 3e$^-$ at 100 kHz to 10e$^-$ at 5 MHz in non-EM mode, 25e$^-$ at 5 MHz with EM gain, and 50e$^-$ at 10 MHz. The quoted CIC value is 0.005 e$^-$/pixel/frame. The EM gain is variable from 1 to 1000 [102].

The ImagEM camera communicates via a protocol known as Camera Link, which is not supported by most computers, and so requires extra hardware. The camera works strictly in frame-transfer mode. The ImagEM boasted a lot of on-board image processing, including image averaging, dark-frame subtraction, and special image enhancement mode called Photon Imaging Mode which supposedly gives better image quality at low light levels, and even increase signal intensity by 5, 13 or 21 times, based on the particular mode used. Only three readout speeds are available, 0.69 MHz, 2.75 MHz, and 11 MHz, with 11 MHz only available in EM mode. The quoted readout noises seem very low, but is oddly dependent on the EM gain: 8e$^-$ at 0.69 MHz in no-EM mode, 17e$^-$ at 2.75 MHz in no-EM mode, and 8e$^-$, 20e$^-$, and 25e$^-$ at 0.69 MHz, 2.75 MHz, and 11 MHz respectively, with an EM gain of 4 (the lowest setting). The same sheet also quoted < 1e$^-$ readout noise for all readout speeds with the maximum EM gain of 1200. No mention of a typical CIC value was given in the data sheet for the camera, although a related technical note seemed to give a value [38, 101].

The iXon communicates with a PCI/PCIe card developed by Andor, as opposed to the more standard protocols used by the other cameras. Like the others cameras, this camera has several image-processing options, and, like the ProEM, the maximum EM gain is 1000. The readout speeds are very similar to the ProEM, and some, but not all, of the quoted readout noises are on par with those quoted by the ProEM: 6 e^- at 1 MHz (no EM), and ranging from 21 e^- at 1 MHz (EM) to 49 e^- at 10 MHz (EM), with intermediate speeds of 3 e^- (32 e^-) and 5 e^- (42 e^-). No noise was given for the 3 MHz no-EM readout mode. While these numbers tend to be a little higher than the equivalent for the ProEM, we found the iXon performance to be much closer to these numbers than the ProEM performance was to its quoted specifications. The quoted CIC value, like the ProEM, is 0.005 e^-/pixel/frame. The iXon is also the only camera of the ones mentioned here that has a 14-bit digitization (16-bit only at 1 MHz readout speeds), while all the rest were 16-bit, although we do not feel this is an important distinction given that all the cameras have modes where the readout noise is larger than a single ADU [105].

The specifications for the Evolve:512 and the Rolera cameras are very similar to those quoted for the ProEM, except the communication is over an IEEE-1394 interface instead of Ethernet. Both the Evolve:512 and Rolera readout speeds were 1.25 MHz and 5 MHz for non-EM modes, and 5 MHz and 10 MHz for EM modes. For those speeds, the quoted Evolve:512 readout noises are 6 e^-, 12 e^-, 32 e^-, and 45 e^-. The Rolera readout noises, in the same order, are 8 e^-, 15 e^-, 40 e^-, and 55 e^-. The EM gains for both are variable with a maximum of 1000. The Evolve:512 specifications also mention several image processing options, including one which is claimed to almost completely remove the CIC, which was quoted at being 0.0045 e^-/pixel/frame, but this turned out to just be a spike removal which would also remove actual photon detections [106]. We do not have a CIC specification for the Rolera, but in communications with a representative for QImaging, we were told that the Rolera operated at a higher temperature than the other cameras, which should, we were told, allow it to have much lower CIC. We did not find this to be the case, as we will show in Section V.9.

All cameras also mentioned thermal noise statistics. Normally, this would be an issue, except all the cameras had on-board coolers. The ImagEM cooler normally can keep the EMCCD cooled to -65°C at room temperature (which is the only value the software we tried would set). The ProEM software allowed a little more variability, and we were able to barely achieve EMCCD temperatures around -80°C at room temperatures, although the software defaults to -70°C, which is what we used for many of the images we took. The iXon and Evolve:512 both claimed to be able to cool to at least -85°C, while the Rolera only cools to -25°C. At these temperatures, the quoted thermal noises are so low (typically less than 0.01 e^-/pixel/second, and closer to 0.001 e^-/pixel/second, but 0.5 e^-/pixel/second for the Rolera) that unless we were to take a multi-second exposure, the CIC would be by far the dominant noise source (except for the Rolera) [105, 106]. Since we plan for our exposures to be less than 50 ms or so, we figured we could safely ignore the thermal noise, but see Section V.7 for a brief discussion of measuring that.

All the cameras (except possibly the Rolera) allowed for restricting the actual readout of the EMCCD to a small region, which allows for a faster frame rate. Reading a full frame at about 10 MHz is limited to about 30 to 40 frames per second, but when restricted to a small region of the EMCCD, frame rates of several hundred per second are achievable.

The cameras, since they all have the same EMCCD, at least had the same quantum efficiency as a function of light wavelength, although the eXcelon coating changed the curve at lower wavelengths on the ProEM than we were interested in. The quantum efficiency for the cameras at 780 nm is between 70% and 80%. As shown in Figure 5.21, the eXcelon coating does reduce fringing at our wavelength. Other elements of our optical setup (including the fused silica cell on our vacuum chamber where we trap atoms) cause etaloning effects, so we hope that a little extra etalon effect from the camera will not be a problem.[6] Furthermore, we will probably be dealing with such low intensities that the presence of fringes will be the least of our worries.

[6] An *etalon* is a transparent plate with reflecting surface. Here, a piece of glass or fused silica acts as an etalon, since a small amount of light reflects off of each surface. This light interferes with the light that did not reflect, and produces intensity fringes. The CCD itself, to some extent acts like an etalon as well, and the eXcelon coating is a sort of anti-reflection coating that prevents this from happening [102].

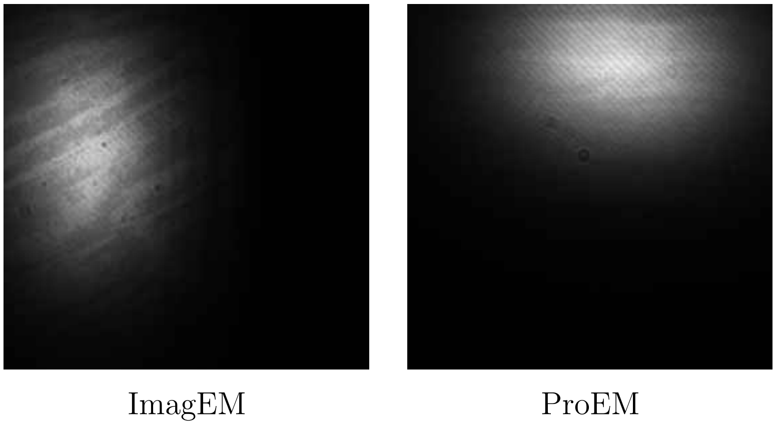


ImagEM ProEM

Figure 5.21. The fringe-reducing effect of the eXcelon coating. Here, we were shining a highly attenuated laser beam directly onto the EMCCD (this is the beam we used to measure quantum efficiency, as described in Section V.10), and took an image with it. The beam was attenuated enough that we could image it at full EM gain with a short exposure, and the image was still noise-dominated (the images shown here are actually the average of 128 sequential exposures, which greatly reduces the noise). The intent was to have the beams illuminate approximately half the EMCCD, so we could have an illuminated region and an unilluminated region. There are three effects here. The round, nearly circular fringes are probably from dust on previous optics. There are also many finely spaced parallel fringes through both images, which we did not bother to track down. Lastly, there is a large, rather splotchy set of fringes in the ImagEM image. We believe these are due to etaloning within the EMCCD itself, and that it is the eXcelon coating on the ProEM EMCCD that prevents them from showing up on that image. This belief is based upon similar images of fringing in the ProEM specification sheet, as we did not investigate the source of the fringes ourselves. The beams have different perpendicular orientations in these images because the EMCCDs in the cameras are rotated differently with respect to the mounting bracket of the camera.

Up front, features such as the eXcelon coating are the most obvious differences between the cameras. The Kinetics mode offered by the ProEM is a variation of the frame transfer mode that both the ProEM and ImagEM use for taking an exposure (see Figure 5.3). In this mode, more than half of the CCD is blocked off. Some versions of the camera have the first two rows exposed, and the next 98 rows masked off (and the user is free to block off the rest). After an exposure, the CCD is vertically shifted so that the exposed area is moved under some mask. If we were to expose 64 rows, and had the rest of the chip masked off, we could repeat this 16 times and get 16 exposures before needing to read the data off the EMCCD. Since we do not need to read the EMCCD between exposures, we can have very rapid exposures. The specification sheet states that this mode is capable of exceeding one million exposures per second, but this is the limit of almost no exposure time (where the entire time is spent shifting) if we are only exposing two rows. Since the fastest vertical shift if just under half a microsecond, if we were exposing 64 rows, each exposure would need to be at least 32 μs, and, if we wanted to prevent smearing, we would want the actual exposure to be longer than just the shift time [102]. Even so, we could achieve a frame rate of about 100 kHz, which is quite impressive, and a feature we might want in a camera.

The Photon Imaging Mode that the ImagEM has turned out to be less interesting. Essentially, the mode is convolution of the image with a hard-edged circle. Put another way, each pixel of the image, instead of being just a pixel, is represented by a hard-edged circle of the same brightness, and these are summed up. The quoted gains of 5, 13, and 21 just give the size of the circles, as 5 is the number of pixels at most a distance of 1 away from a given pixel, 13 is the number of pixels at most a distance of 2 away, and 21 is the number of pixels at most a distance of 2.5 away. These shapes are illustrated in Figure 5.22. Thus, an apparent gain in brightness comes from essentially averaging over a wider area of pixels, effectively getting 5, 13, or 21 times as much light, but at a cost of resolution. This may be useful to some, but since we have plenty of time to post-process our images, and are quite capable of performing this convolution (or any other that may be more useful to us) ourselves, this feature is not that interesting to us. Some of the ImagEM

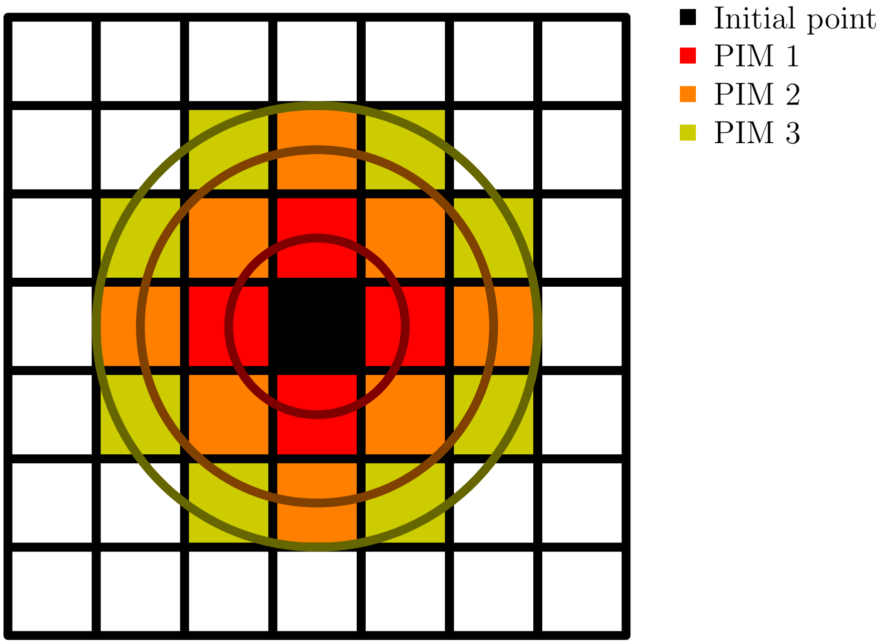
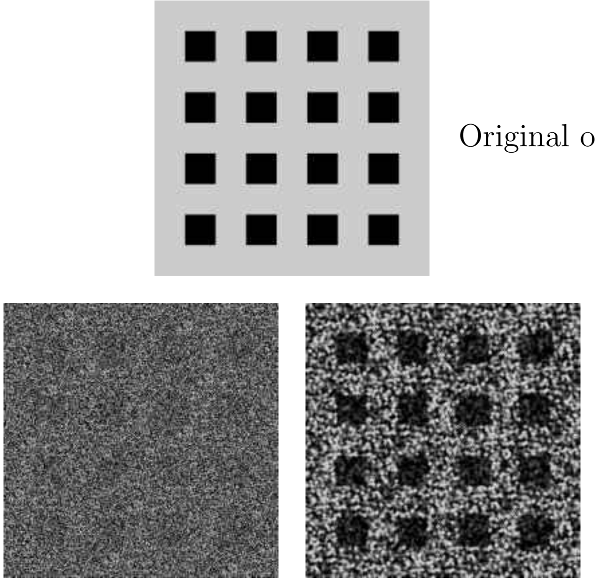


Figure 5.22. Photon Imaging Mode circles. These three circles have radii of 1, 2, and 2.5 pixels. The shaded pixels inside those circles represent the shapes that are convolved with images in the three Photon Imaging Modes supported by the Hamamatsu ImagEM cameras. Note that the number of pixels within the three circles are 5, 13, and 21, respectively, which are the quoted gain numbers for the three Photon Imaging Modes.

documentation have some sample images that show a huge improvement in clarity when imaging a very faint object [38]. Figure 5.23 shows a simulation of viewing a very faint pattern with an ImagEM camera using Photon Imaging Mode 2, which illustrates how the convolution can produce a remarkable enhancement of visual contrast. When generating this particular figure, however, we found that there was a rather small window of contrast for which there was a marked improvement (but we did manage to get this figure to be very similar to one involving actual camera images distributed by Hamamatsu). If the image was much brighter, the Photon Imaging Mode convolution just blurred the image, and if the image was much dimmer, it was too dim to make out even with Photon Imaging Mode.

The Evolve:512 offered a CIC-removal processing mode called Background Event



Original object

Simluated image of object

Simulated image of object, with Photon Imaging Mode

Figure 5.23. A simulation of Photon Imaging Mode. Here we show a pattern of squares, and a simulation of what that pattern would look like if it were being imaged by a Hamamatsu ImagEM camera with an average illumination of under one photon per pixel per exposure. Under the specific conditions of this simulation, performing the Photon Imaging Mode convolution greatly enhances the visual clarity of the image. However, when generating these images, we have approximately matched actual camera images distributed by Hamamatsu, and found that we had to fine-tune the illumination to get that enhancement [38].

Reduction Technology (BERT) [106]. In this mode, every pixel is compared with the surrounding pixels 9 pixels, and, if it was above the median by a certain threshold, it was replaced with the median value. While this is a nice way to clean up the CIC in an image, it irreversibly alters the image data and could cost us a single-photon signal, and so we decided this feature would not be useful to us. If we felt we needed this ability, we could always do it in post-processing.

A less-apparent difference between the cameras is in the way the specifications are quoted. For example, the readout noise quoted for the ImagEM camera is gain dependent, whereas no such indication is given for the other cameras. The reason is that the ProEM number is the actual readout noise, whereas the ImagEM is the *effective* readout noise. Note that in any representation of the signal-to-noise ratio, such as Equation (V.4), or the more formally derived versions, Equation (A.14) and Equation (A.15), and in the histograms we are using (Equation (A.56), but also shown earlier in this section), the parameter that matters is the ratio of the EM gain to the readout noise. The readout noises shown for the ImagEM are therefore the actual readout noises divided by the EM gain. For example, the actual typical readout noise at 11 MHz is 100 e^-. At the lowest EM gain for the C9100–13 ($G = 4$), the effective readout noise is $100\ \mathrm{e}^-/4 = 25\ \mathrm{e}^-$, and at the maximum EM gain $G = 1200$, the effective readout noise is $100\ \mathrm{e}^-/1200 = 0.08\bar{3}\ \mathrm{e}^- < 1\ \mathrm{e}^-$, which is where the quoted values come from. Our own measurements for the readout noise agree with this value, if we infer the readout stage scaling by assuming the maximum EM gain is really 1200. So, if we really wish to compare the ProEM and ImagEM readout noises, we must first multiply the ImagEM readout noises by the EM gain used.

Over the course of the next few sections, we will discuss some of the values we actually measured using our curve-fitting techniques. In particular, we will see that the quoted noise values often needed adjusting in order to be properly compared between cameras, justifying the work we did to measure the values ourselves.

Comparing EMCCD CIC and Thermal Noise Values

Our first attempts to measure CIC noise were to take a series of dark frames, compute the variance of the frames, and use that to determine the CIC by using Equation (V.2). We also tried several exposure lengths, and did a linear fit to the variance, which should also give us the thermal noise, since thermal noise is just an added contribution to variance that increases with time. We realized this method was failing us when we found that, for several sets, the thermal noise appeared negative. There were two basic reasons for this. First of all, variances are sensitive to outlying points. In our case, some of the images have cosmic rays in them. Very few pixels have these rays, but those pixels often have large values, and so raise the total variance by a noticeable amount. The second reason is that the CIC and thermal noise are pretty small contributions to the variance. A large gain can make the readout noise irrelevant to the total signal-to-noise ratio, but in a dark frame, the signal is zero. For typical camera parameters where the CIC is on the order of a percent, the gain is on the order of a thousand, and the readout noise is on the order of ten to a hundred electrons, the CIC contribution to the variance is still of the same order of magnitude as the readout noise. Cosmic rays were able to affect the variance by enough of that ratio to mess up the fits.

We considered dealing with this problem by taking many more images, filtering out the cosmic rays, and then using the same variance method. However, we realized we could use the histogram fits that we now use instead. We chose the histogram methods because it allows us to bypass the effects of cosmic rays (which tend to appear as small unfittable spikes in the histograms), provides us with the ability to model similar cameras would do under different imaging circumstances, and is a little less sensitive to random variations (like cosmic rays, but also just deviations between images). It also gave us some intuition on how to compare cameras by just looking at their histograms.

Once we developed this method and applied it to the sample images we took with demo cameras, we very quickly noticed a large difference in CIC values, both between the ProEM and ImagEM cameras, and between our measured values for the ProEM camera and the quoted value. At the time we were performing these tests, we only had data from ImagEM and ProEM cameras. Typical best values for the ImagEM cameras were on the order of a few percent, while the ProEM measurements were, at best, over five percent, and often over ten percent, quite different than the half-percent quoted in the specification sheet. The discrepancy between cameras was quite noticeable in the histograms, too. As seen in Figure 5.6, the CIC tail is only visible in the ProEM histogram, and in the semi-log plots, Figure 5.7, Figure 5.14, and Figure 5.15, the ProEM can be seen to have a much larger vertical offset to the CIC tail than the ImagEM. These figures represent some of the best CIC measurements for both camera models, and so are a good indication of the difference between the values.

We found out, through several communications with the manufacturers, that the ProEM CIC values quoted are not the mean CIC value that we thought they were. The quoted CIC value is defined as the fraction of pixels with values over 5 standard deviations from the mean value. This measure is apparently standard in the industry [40], but seems quite strange to us. Using this method, we can compute values close to the quoted CIC values for the ProEM camera, but we note that this number is *not* the actual mean CIC value, and is often times almost an order of magnitude smaller. In fact, it is not even a direct measurement of the mean CIC value, as it depends on the particular EM gain and readout noise values. We show this dependency in Figure 5.24, where we plot curves of this quoted CIC value for a fixed CIC and readout noise typical of a ProEM camera, with a varying EM gain. Here, we see that at the maximum gains of the cameras we were using, the gain-dependence of this measure of CIC is fairly weak, so it is, in a sense, a measure of CIC, but needs a correction to give the actual mean CIC value.

As an aside, we also made brief attempts to measure thermal noise. Thermal noise, from a histogram standpoint, is indistinguishable from CIC noise, as both are just Poissonian noise that gets added to the signal. As previously mentioned, they tend to focus in different physical depths of the CCD, and actually can be marginally distinguished by, for instance, the clocking methods, but, for our purposes, they are indistinguishable. Basically, we can compute the CIC for various exposure lengths, and perform a linear fit to that CIC as a function of exposure length. The time $= 0$ intercept is the true CIC value, and the slope with respect to time is the thermal noise, in electrons per second.

We were only interested in short-exposure statistics, since we plan to take mostly short exposures, so only rarely did we take exposures long enough to see thermal effects. For one ImagEM camera, we found tentative values ranging between 0.002 e^{-}/pixel/second to 0.007 e^{-}/pixel/second, depending on the imaging mode. Again, these typically had short exposure times (at most about 2.5 s, but typically under a second), and so the increase in apparent CIC over the $t = 0$ value was rather small compared to our estimated accuracy, which is why these values are

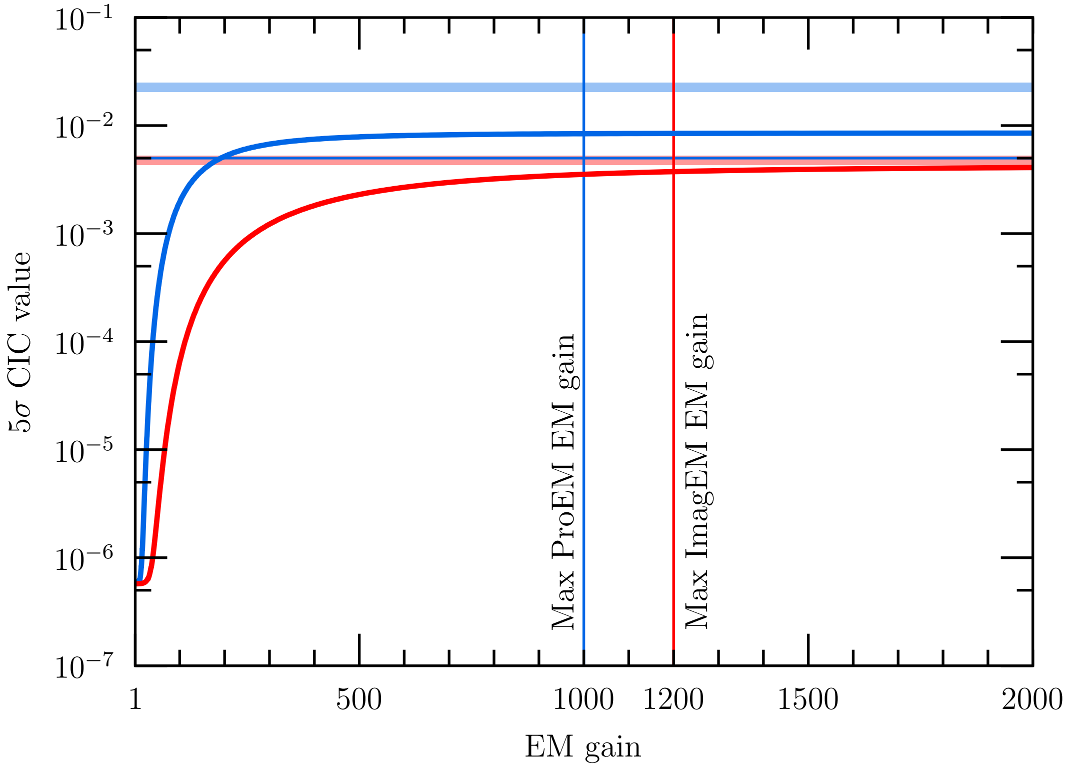


Figure 5.24. Comparison of 5σ CIC measurements with mean CIC. We took one ProEM data set and one ImagEM data set, and, using the noise parameters we measured for them, used our model to generate histograms for a wide range of EM gains. We then computed the 5σ CIC value for those histograms, and plotted the results, using blue for the ProEM curves and red for the ImagEM curves. The two vertical lines show the maximum EM gains of the cameras (which Princeton Instruments uses for quoting a CIC value), and for which we have actual camera histograms. We verified that at those gains, the values shown on the plot closely match what we measured from the actual camera histograms (within about 10%). The horizontal bands of color show our measured CIC value that we used to generate the plots, and the horizontal line shows the value that Princeton Instruments quotes for the ProEM cameras. We see that for the maximum EM gains, the 5σ value is not strongly dependent on the actual EM gain, and so could be used to deduce the mean CIC value, but the two values are nearly an order of magnitude different. Although not shown here, we also note that the amount of EM leakage versus actual CIC changes the shape of the curves as well.

tentative. These values are slightly better than the quoted typical value for this camera (0.01 e$^-$/pixel/second at -60°C) [101].

All the relatively long exposures we took with ProEM cameras were done with the first ProEM demo camera we had. Unfortunately, this camera had large noise problems, with an obvious shading problem that was different between images (and hence did not subtract out well), and a few other repeatability problems.[7] The result was that we could not fit the histograms well with our model, and the results tended to change a lot even for small changes (10 ms to 20 ms exposure times, for instance). While we determined that the noise levels were abnormally large for this camera, the oscillations within the noise measurements were too large to get a decent estimate of the thermal noise. The ProEM specification sheet quotes a typical thermal noise of 0.001 e$^-$/pixel/second and a maximum thermal noise of 0.02 e$^-$/pixel/second, both at -70°C, but we do not know how our measurement would compare with these values, or whether these values are using the 5σ measurement method or are actual mean values [102].

Comparing EMCCD Readout Noise Values

The curve-fitting method gives us a readout noise measured in ADU. More importantly, it gives us the ratio of EM gain to readout noise, which is more important to us than the actual readout noise. However, the actual readout noise matters in the low-EM limit, and to anybody attempting to improve the performance of a camera one noise-source at a time. That, plus a few odd anomalies we discovered, warrants a quick discussion about readout noise.

In order to compare readout noises between cameras and even between different modes of the camera, we need to veer away from measuring readout noise in ADU. The ProEM cameras all came with a specification sheet that actually gave the ADU-to-e$^-$ conversions for the different gain modes of the cameras, the ImagEM cameras

[7] Demo cameras should be treated as slightly suspect. They are shipped often and perhaps not always treated with the best of care by people trying them out, and thus may not be in peak condition. One camera we had as a demo actually had a loose screw rattling around within it.

did not, and we were only able to get such conversions for some of the dark frame sets we got for other cameras. However, all the cameras claimed to have calibrated EM gains. If we assume those gains are accurate, we can use our measurement of EM gain in ADU (as the average value of an amplified single electron in ADU), and use the actual EM gain to infer the ADU-to-e^- conversion. Doing this on the ProEM camera, we typically found conversion factors within 5% of the values given by Princeton Instruments. For the ImagEM cameras, this gave readout noises within about 5% of the quoted typical readout noise (once we multiplied the quoted value by the EM gain, as described above). This gives use reasonable cause to believe we can accurately compare readout noises between camera models, and very strong evidence that we can compare readout noises between different imaging settings on the ProEM cameras.

For the ImagEM camera, we consistently found readout noises very close to the quoted typical value [101]. For the ProEM cameras, though, we consistently found readout noises almost twice the quoted value [102]. Princeton Instruments has put a lot of effort into reducing their readout noise, and quote much lower readout noises, on the order of a factor of two lower for comparable readout speeds, so it seemed important that we be able to reproduce their quoted noise measurement.

After several discussions with people at Princeton Instruments, we found two major discrepancies. The trick to finding these discrepancies lies in the different methods we use to measure readout noise. We use our curve-fitting method, while Princeton Instruments takes a series of images (with EM gain either off or at the lowest setting), and then takes the standard deviation of the difference of the last two images in the series. This deviation is then divided by the square root of two, and that yields the quoted value [40]. Taking a series of images allows various start-of-sequence effects to settle out before the measurement takes place, taking a difference of two images removes any constant shading, and dividing by the square root of two properly accounts for an effective doubling of the variance by adding two independent noise sources (the two images). There is no reason why our curve-fits should not work with images that have very little EM gain. The histogram then should basically just be a convolution of readout noise with a zero-electron peak (plus a small amount of CIC), which our routine can fit. However, since there is no CIC tail, the curve-fit cannot tell whether there is basically no EM gain or no CIC, or both. Thus, fitting a histogram with very low (or no) EM gain typically fits the curve decently, and gives good readout noise values, but tends to give bad CIC and EM gain values. We found that we agreed reasonable closely when we used both our methods on the same set of images (which required using low EM gain or no EM gain).

The first reason for the discrepancy, we found, was that the readout noise was highly dependent on the EM gain. The CIC tail, which is effectively non-existent at low EM gain, increases the standard deviation of an image at high EM gain, and so the image-difference method cannot be used to see this effect without some other way to correct for the CIC tail. Our histogram method, however, shows the problem quite clearly. In Figure 5.14, we can clearly see that the ProEM has less readout noise than the ImagEM camera, by comparing the width of the 0-count Gaussian bumps (which are scaled into electrons). The ImagEM has very nearly the exact same readout noise from the lowest to the highest EM gain. Ignoring the odd bumps on the side of the ProEM histograms for now, we clearly see three different widths for the readout noise on the three ProEM histograms. The ProEM has more readout noise in the EM modes than in the non-EM modes (as mentioned in the specification sheet), so the difference between the no-EM and EM modes for the ProEM are expected. However, the remaining two ProEM histograms were taken in the exact same imaging mode, and only the EM gain was changed. Here, we clearly see a large increase in readout noise that the ImagEM did not have. While this increase does not raise the readout noise of the ProEM to the same amount as the ImagEM, it certainly reduces the low-readout-noise advantage that the ProEM has over the ImagEM without EM gain.

The second reason for the discrepancy has to do with how we took the images. The way Princeton Instruments measured readout noise involved throwing out the first few images of the series of images they took. In the camera we tried this on, we found the first image tended to have slightly more image shading and larger differences between adjacent columns (see Section V.11 for more on this). These settle down after the first image or so, but create an abnormally large readout noise for that first image.[8] Furthermore, we were initially taking images in such a way that each of the 128 images we thought were a single sequence were, in fact, each the first image of a sequence. The ProEM actually has two readout amplifiers, and those two amplifiers tend to have a random relative offset compared to each other [102]. As discussed in Section V.11, that relative offset normally shifts around by an amount much smaller than the readout noise, and so the effect is negligible. In the first image of a sequence, however, that offset (and the extra readout noise) could be comparable to the readout noise, since it drifts by that much from first image to first image, and adds to the apparent readout noise even more. The extra shading also increase the readout noise, but this is usually (but not always) a lesser effect.

This effect can be seen in Figure 5.14. The data in this figure was taken from a single series of images, and the strange bumps on the left side of two of the ProEM histograms are solely from the first images of the series, which, in both these series, had a different offset than the rest of the series. It is not so obvious from the figure that the relative offsets of the columns are different, but it is obvious that the average offset was definitely different. Also, note that the amount by which they differed from the rest of their series is different between the two images.

We can also see the effect in Figure 5.25. In this figure, we take three series of 128 images each, one from an ImagEM, one from a single series of images from a ProEM, and one from the ProEM where each image was the first of a sequence. All the readout noises are given in electrons, for comparison between cameras (and computed as described in the text for Figure 5.14). The thick horizontal stripes show the quoted typical readout noise values for the two cameras. The dashed lines show the readout noise measured by curve-fitting the aggregate histogram from the 128 images. The solid lines show the readout noise measured using the standard deviation of the difference of two sequential images. The ImagEM has the most consistent readout noise, just a little below the quoted value. The corrected ProEM

[8] We refer to this effect as the *first-of-sequence problem* in this chapter. The increase in shading appears to be from a slight vertical gradient on the images, probably from the readout amplifier offsets slowly settling to their steady-state value during the image readout.

series (where each image is from the same sequence) has slightly more variation, but is right at the quoted value, and agrees with our curve-fitting method. Note that the first and last point of that curve, which are the two that used the first image of the sequence, seem to have a larger readout noise than the rest. The third curve shows the sort of results we were originally getting, when each of the images was the first of a sequence. It jumps around a very large amount, depending on what the relative readout offsets happened to be for the corresponding pair of images, and how much extra shading there was. Every now and then, the offsets happen to be small (and, typically, the extra shading is also small), and the measured readout noise is close to both the quoted value and the value we measured once we figured out how to deal with the problem. More often, though, the readout noise seems much larger when using first-of-sequence images.

Our general conclusion is that while the ProEM does have less readout noise than the ImagEM, the difference is greatly reduced from the quoted specifications if you turn the EM gain up high, and is reduced even more if you are not careful about whether your images are the first of a sequence or not. While the latter effect can be prevented with some extra work from the user, the former seems to be a problem the user of the camera will have to live with.

Comparing EMCCDs Through Histograms

We now demonstrate our results from using our histogram model on many different settings on the several different ProEM or ImagEM EMCCD demo cameras we tried. We performed curve-fits on many different camera modes, and ranked the modes based on the CIC noise values in Figure 5.26 and the EM gain (relative to the readout noise) in Figure 5.27. The points are color-coded and shape-coded to help with point identification. Red, orange, and yellow refer to Hamamatsu ImagEM cameras, with readout speeds of 11 MHz, 2.75 MHz, and 0.69 MHz, respectively. Blue and green refer to Princeton Instruments ProEM cameras, with readout speeds of 10 MHz and 5 MHz, respectively. Some of the sets include cameras both with (the ProEM BK model) and without (the ProEM B model) Kinetics

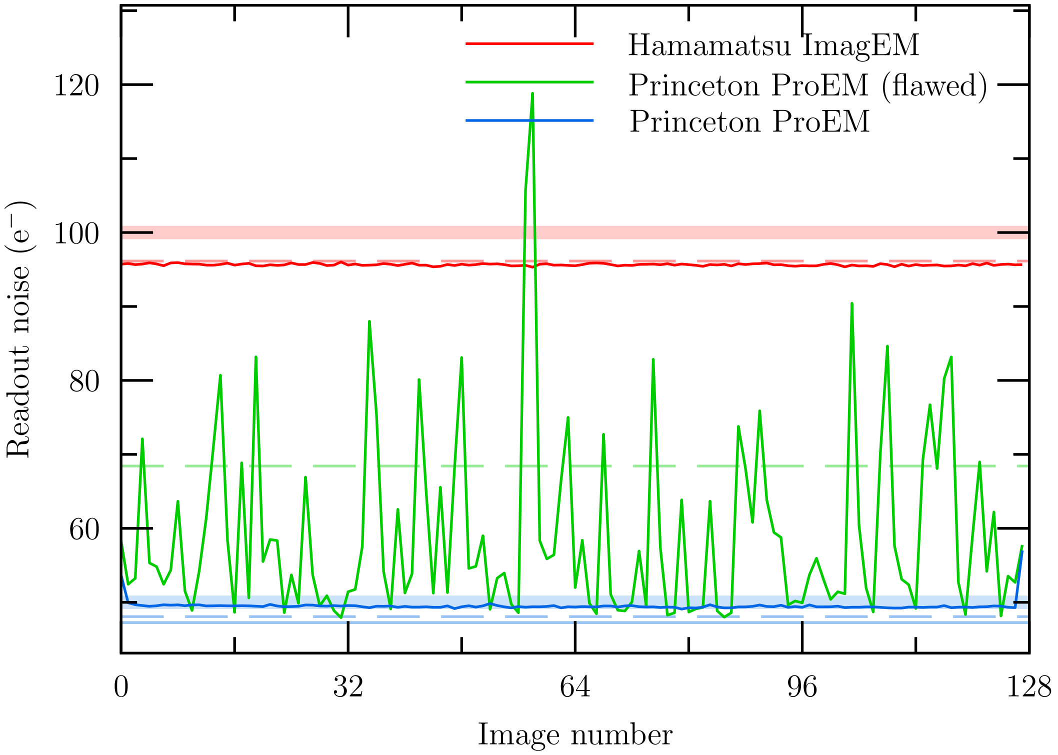


Figure 5.25. A comparison of readout noise measurements, in a low-EM-gain mode. All the values were converted to electrons as described in the text corresponding to Figure 5.14.

mode, which is why the legend has "BK/B" for some sets. Dark blue and dark green are the ProEM same cameras (and speeds), but using some method to work around the first-of-sequence problem described in Section V.8 and Section V.11. The pink color refers to an Andor iXon camera, with a readout speed of 10 MHz. The dark and light tan colors indicate Photometrics Evolve:512 cameras (10 MHz and 5 MHz, respectively), while the light blue squares and light green squares represent the QImaging Rolera Thunder data sets. All images used EM modes, which rules out the 1 MHz and 100 kHz readout speeds on the ProEM. Gray and black refer to the first ProEM camera we had, which turned out to have very bad shading problems and abnormally large noise values, which resulted in poor curve-fits and poor performance, as can be seen here. The shapes of the points on the graph (circles, triangles, and squares) refer to which particular camera was used within a color group. Hollow shapes refer to images taken with the fastest vertical clock shift (450 ns) available for the ProEM, a mode which should have less CIC than other speeds, although we did not see a strong difference. All other ProEM images were taken at 600 ns vertical clock shift, with the exception of two points, which were taken at 2000 ns and 5000 ns vertical shift times. The fourth- and eleventh-highest CIC values in Figure 5.26 (third- and fourth-highest not counting the first ProEM demo camera) used a workaround for the first-of-sequence problem that made the images substantially worse. We have relatively few points for the iXon, Evolve:512, and Rolera, as we were limited to the data sets we had for those cameras.

There are some interesting trends in these figures. First of all, both plots show some sorting based on readout speed. According to Hamamatsu's EMCCD technical note, the fastest readout mode is optimized for CIC performance, while the other two modes are optimized for low thermal noise [38]. We can easily see that in Figure 5.26, where the fastest readout modes have much less CIC then the two slower readout modes, but even the two slower readout modes seem to show a slight difference in CIC. While the ProEM modes are supposedly also optimized for CIC at faster speeds, and thermal noise at slower speeds, we do not see as much sorting based on imaging mode here [40]. Of particular note, the 450 ns vertical shift mode, represented on the graph by hollow points, was supposedly the most optimized for CIC, and the 600 ns vertical shift mode, which is all but two of the solid ProEM points, was supposedly more of a compromise between CIC and thermal noise, but we see little obvious difference between these. The only obvious note is the two highest-CIC modes are the only two points on here which were taken with the long vertical shifts (2000 ns and 5000 ns). These two points were almost certainly optimized more for thermal noise. The best ImagEM, iXon, and Evolve:512 modes have almost an order of magnitude better CIC than the best of the ProEM modes, and the Rolera does not fare well on this metric.

The sorting also occurs when looking at readout noise in Figure 5.27. The ImagEM sets are again sorted by readout speed, with the slowest speeds having the lowest noise (and thus the largest relative EM gain), as expected [38]. With an

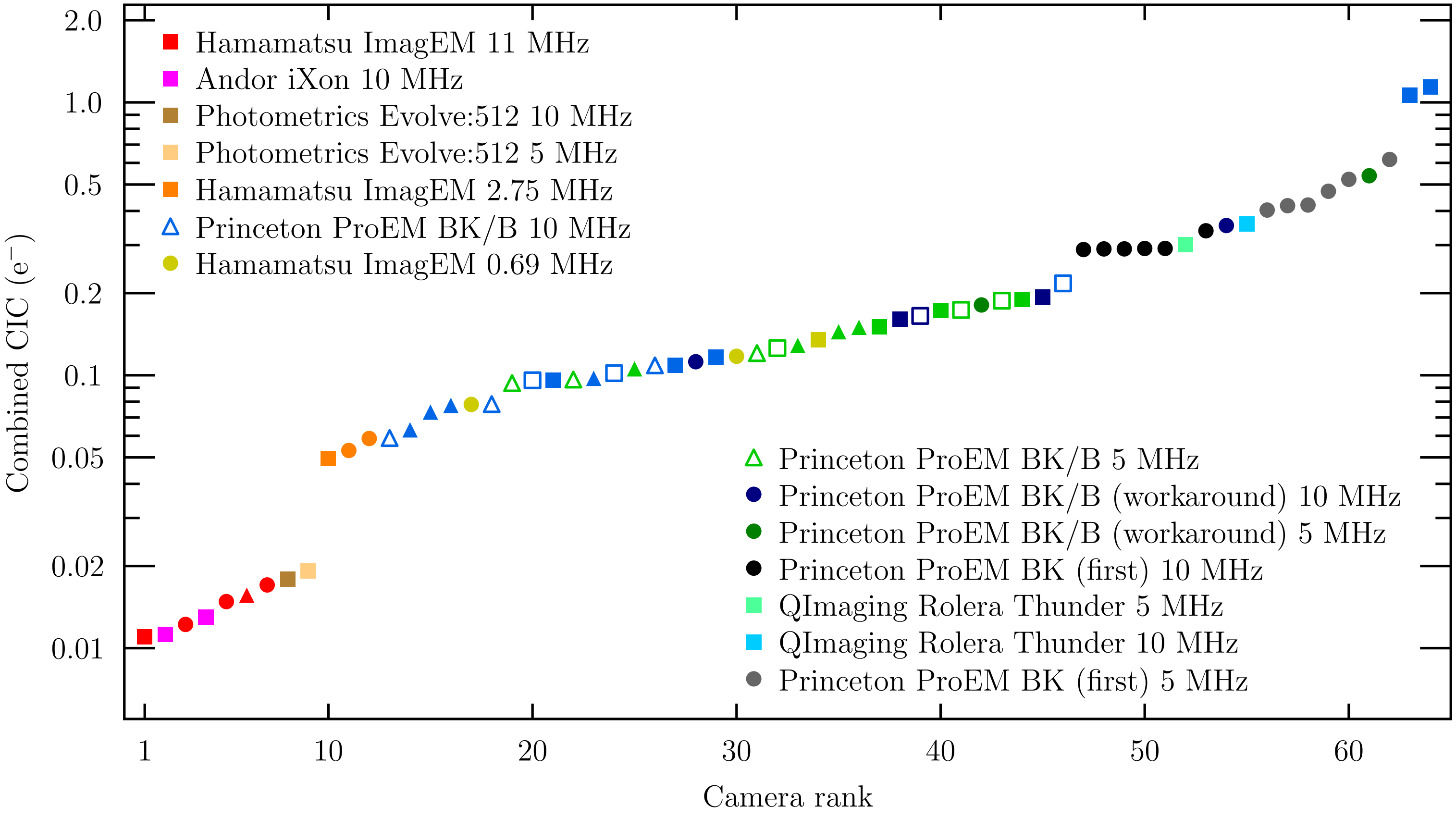


Figure 5.26. Our measured CIC values for various cameras in various imaging modes, ranked from lowest to highest. We do not have error bars on this plot, but a decent estimate for errors may be found by comparing measured values for modes that should have very similar CIC values. The color scheme and shape scheme is described in the text.

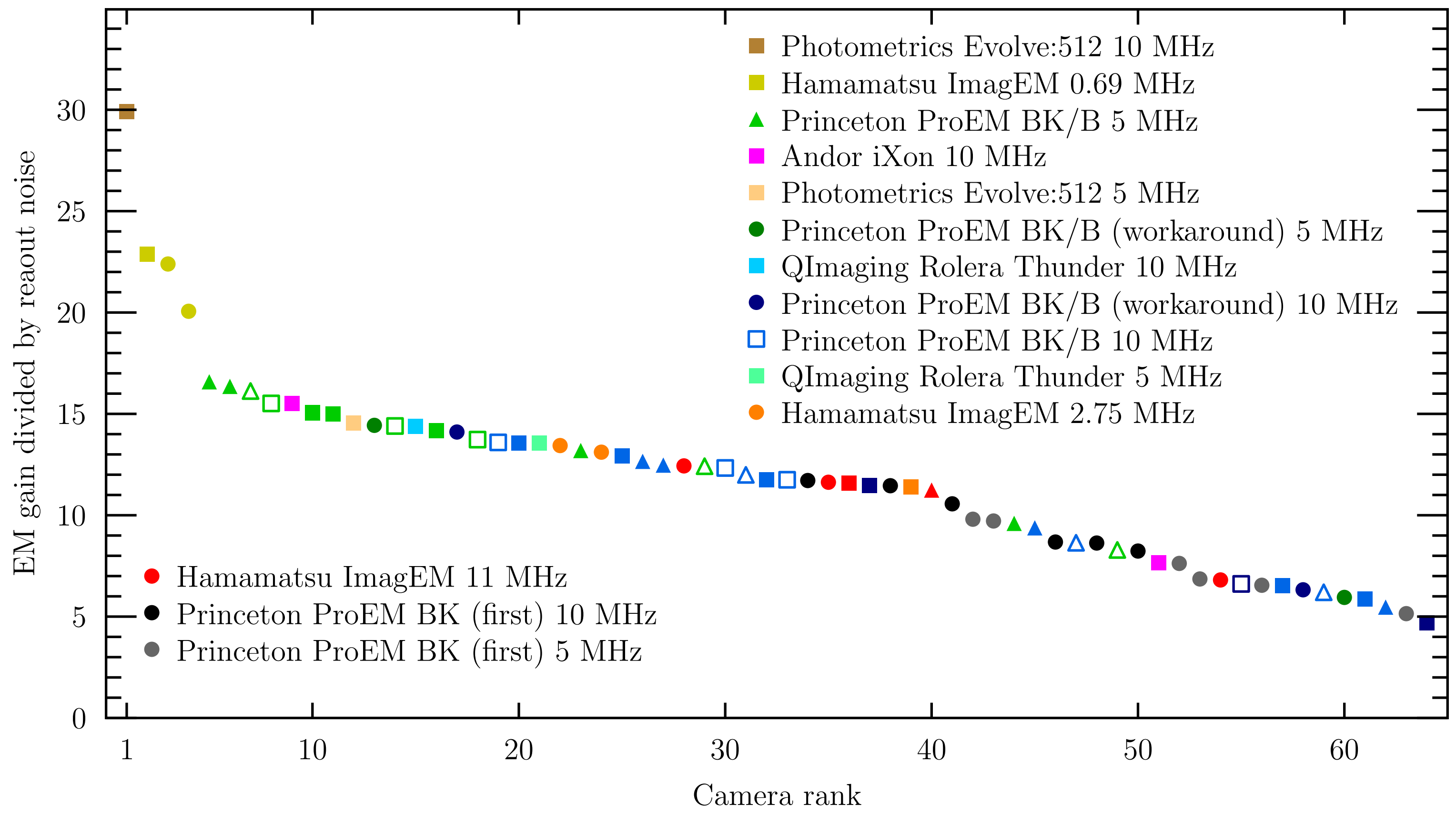


Figure 5.27. Our measured values for EM gain divided by readout noise for various cameras in various imaging modes, ranked from highest to lowest. We do not have error bars on this plot, but a decent estimate for errors may be found by comparing measured values for modes that should have very similar values. The color scheme and shape scheme is described in the text.

EM gain of 1200, the 11 MHz modes (typical readout noise of 100 e$^-$) should be at about 1200/100 = 12, the 2.75 MHz modes (typical readout noise of 80 e$^-$) should be at about 1200/80 = 15, and the 0.69 MHz modes (typical readout noise of 32 e$^-$) should be at about 1200/32 = 37.5. The 11 MHz modes hit this value quite closely, while the slower modes seem to have more noise (or less EM gain, but the noise seems more likely) than the quoted values [101]. There are also a few outlying points (for both the ImagEM and the ProEM) with apparently much less relative EM gain than the others; these actually were taken with approximately half the maximum EM gain, and so the lower apparent EM gain is real but not useful in ranking the cameras. The ProEM modes are not quite as sorted as the ImagEM modes, but there is a trend for the slower readout speeds to perform better than the faster readout speeds. From the specification sheet, with the EM gain set at 1000, we would expect the 5 MHz modes (typical readout noise of 25 e$^-$) to be at 1000/25 = 40 and the 10 MHz modes (typical readout noise of 50 e$^-$) to be at 1000/50 = 20, but these are quite a bit off [102]. The main reason for this discrepancy is that the readout noise nearly doubles as the EM gain is increased. This problem is noted in Section V.8. The other problem mentioned in that section, where the first image of a series causes high readout noise measurements, is corrected for in some of the data points in Figure 5.27, and seems not to have been too much of a problem in others (since others seem to have slightly less noise than sets that had this problem corrected). Again, the first ProEM demo is considered to have abnormally high noise and bad curve-fits that make these measurements particularly suspect. Most of the ProEM sets here are affected by the first-in-sequence problem described in Section V.8, with the exceptions being the colors labeled “workaround” in the legend. While correcting for this problem helped some data sets, they are bested by some other data sets with the problem, where, presumably, the problem just happened to not be much of an issue.

The clear winner in Figure 5.27 is the 10 MHz Evolve:512 point, which is odd, because we would have thought the 5 MHz Evolve:512 point would have less readout noise than the 10 MHz Evolve:512 point. However, there are several odd things about this point, which we discuss in Section V.11. With the exception of the anomalous Evolve:512 point, the slowest ImagEM settings are a clear winner here, but the readout speed of 0.69 MHz may be too slow for us. The next best is the 5 MHz readout speed of the ProEM cameras, which, despite being a faster readout speed than the 2.75 MHz ImagEM modes, still has less readout noise. The ProEM cameras have less EM gain (1000) than the ImagEM cameras (1200), so this is entirely due to Princeton Instruments efforts to get a very low readout noise. The iXon, Evolve:512 (5 MHz), and Rolera rank quite closely with the best ProEM points.

The iXon specifications are very similar to the ProEM. For the 10 MHz readout mode used in the figure, the typical readout noise is also about 50 e^-, and the maximum EM gain is 1000. Thus, we would expect a ranking of about $1000/50 = 20$, but instead we see it is closer to 15 (and the second point, with an EM gain of 500, is about half that). We have electrons per ADU values for the iXon camera used in the figure, given to us by an Andor representative. If we choose to trust those, then the readout noise for the camera is quite close to the typical value, but the EM gain used is closer to 820 than 1000. We do not know where the discrepancy really is, but would like to point out that whatever it is, the iXon seems to be about tied for the best EM gain to readout noise ratio of any camera mode with a readout rate of 1 MHz or more, with the possible exception of the first-ranking point.

The top-ranking point in Figure 5.27 ranks quite a bit better than any other camera mode. This is a little odd, because that point is the Evolve:512 10 MHz mode, and it substantially outranks the 5 MHz mode from the same camera. In the other cameras, and in all the specification sheets, we see the readout noise *increases* with readout speed, but we seem to see the opposite effect happening here. We have a few data sets from this camera at low EM gains, and, comparing with these, we suspect this mode has an anomalously high EM gain of over 2000. Since the maximum gain of the camera is quoted as 1000, we suspect this is an EM gain calibration error. We will discuss our thoughts on this point more in Section V.11.

Finally, we notice that the Rolera did not perform at all well on CIC measurements in Figure 5.26. As mentioned earlier, the Rolera does not run at a very cool temperature (-25°C) compared to the other cameras. While this makes the thermal noise substantially higher, a representative for QImaging gave us an argument that the higher temperature allowed the Rolera to achieve much lower CIC than most cameras. Our CIC measurement actually detects the sum of CIC and thermal noise, but we have been calling it a CIC measurement because in all cases, except possibly this one, the low thermal noise of the cameras coupled with the short exposure times used pretty much guaranteed that the measured quantity was almost entirely CIC. However, we have a decent argument that the "CIC" noise seen in Figure 5.26 for the Rolera is mostly CIC, and not high thermal noise. We did not take the images ourselves; rather, a representative took them for us and sent us the files for analysis. We instructed the representative to use the shortest possible exposure time, and the camera settings that were sent along with the files suggest that request was followed. The exposure time for the images appeared to be 0 ms. Thus, we consider it reasonably likely that the only time for thermal noise to accumulate is during the vertical clocking and readout time. The readout time with a 10 MHz readout rate should be approximately half the readout time with a 5 MHz rate.[9] Thus, if the majority of the measured noise were thermal noise, the 5 MHz readout rate should show approximately twice the noise as the 10 MHz. This is not the case. In fact, it is reversed, and the 5 MHz readout has slightly less noise than the 10 MHz. This suggests to us that the noise we have measured is indeed mostly CIC. As we have mentioned previously, the quoted thermal noise for the Rolera is 0.5 e^-/pixel/second. At a readout speed of 5 MHz, reading out a 512-by-512 array would take about 50 ms, plus a little for vertical shifts (including the initial wipe, and shifting the exposed area under the mask). Even if the total exposure time were on the order of 100 ms, we would expect the thermal noise to be on the order of 0.05 e^-. Our measured value is on the order of 0.3 e^-, an order of magnitude larger than what we would expect from the thermal noise. These two reasons are why we think we have actually measured CIC and not thermal noise, but it is not really relevant. Regardless of whether the noise is mostly CIC or thermal, it is a

[9] It would not be exact because the vertical shifts between reading out rows, and for wiping the chip before the exposure, add some time to the exposure that does not necessarily scale inversely with the readout speed, but the majority of the time would be spend reading cells.

minimum which will only increase with exposure time, and so we can still think of the measurement as a minimum CIC-like noise value.

From the CIC rankings in Figure 5.26, we see that the fast ImagEM modes have some of the best rankings. From the EM gain (relative to readout noise) rankings in Figure 5.27, we see that the slow ImagEM modes have some of the best rankings, with several of the ProEM modes in second. The question we would then like to ask is, which ranking is more important? Intuitively, we might expect, since the CIC ranking shows an order-of-magnitude improvement in one type of noise, and the EM gain ranking shows less than a factor of two improvement, that the CIC difference might be more important. Furthermore, since the vast majority of the camera modes have EM gains over 10 times the readout noise, we might expect that having a mean value of 10 standard deviations outside a Gaussian is not a great deal more distinguishable than having a mean value of 20 standard deviations outside a Gaussian, since Gaussians have pretty much no overlap with anything that far away from their centers. Also, since the CIC values mostly less than unity, and the EM gains are large (compared to the readout noise), we might expect the CIC to be most important whenever the expected number of received photons is on the order of unity or less, since a change between 0 and 1 is already easily distinguishable thanks to the large EM gain, but such a change is about as likely to be false (CIC) as not (a photon). Likewise, we might expect that if the mean photon number is large compared to the CIC, then having a lower readout noise or larger EM gain might be more important because that would help lower the overlap between 10 photons and 15 photons, whereas such a difference would be almost impossible from CIC alone. We will now outline a way to quantitatively combine these rankings in a way that is similar to our intuitions.

The basic idea is to assume we have taken an image, and wish to know the amount of light that a given pixel has received. By looking at the image alone, we cannot tell the difference between an electron from CIC noise and an electron from a photon hitting the detector. Thus, a single image cannot tell us with certainty how many photons hit the detector. However, knowing the relative likelihoods that there was an electron, and the likelihood that any electron was actually from a photon, we can compute the probability that a photon was received. We are not as interested in that probability as we are in the spread of the probability; if the images from the camera tend to give sharply peaked probability distributions for the number of photons received, then we have a good measurement of intensity, but if the images tend to give broad probability distributions, then that is hardly any better than random guessing. The entropy of a probability distribution is a measure of the spread of the distribution. A large entropy value means a wide range of options are all likely, whereas a small entropy value means only relatively few options are likely. Thus, we can look at the probability distributions we would expect from such a measurement, and the ones with the smallest entropies indicate the best cameras.

More specifically, we first assume there is a known distribution of incident photons. Technically speaking, we should assume a known distribution of *absorbed* photons, which is different than the distribution of incident photons, since the quantum efficiency of the EMCCDs is less than unity (not all photons are absorbed and produce an electron). Since all the cameras we tested should have the same quantum efficiency (except for hopefully small camera-to-camera variations), we can safely assume the same absorbed distribution for all the cameras. Furthermore, since incident photons usually have a Poissonian distribution, and the random absorptions of photons will result in another Poissonian distribution (as shown in Section B.4), we will assume the electrons from absorbed photons have a Poissonian distribution with a certain mean value (that is technically the product of the incident mean and the quantum efficiency). For simplicity, we will refer to the number of photons per pixel, but actually be referring to the mean value of the Poissonian distribution of absorbed photons.

Once we know the distribution of incoming photons, we can convolve that with a Poissonian distribution for the CIC (and maybe thermal noise). The mean value of the CIC noise is taken from our histogram fits for a particular camera mode, and then use our model to generate a theoretical histogram for that camera mode (using the other parameters of the fit), using this new Poissonian distribution as the input.[10]

We now have a theoretical probability distribution for a pixel of an image taken with a real camera using a particular setting. Since we can also generate histograms for theoretical cases where we know *exactly* how many electrons started in the pixel, we can use these to generate a reverse histogram that tells us, given a final pixel value, what the probability distribution for the initial number of electrons is. Since we also know the distributions for the number of photons and the CIC, we can then compute the probability distribution for the number of absorbed photons for that pixel. Finally, we compute some measure of entropy of that distribution (for a fixed pixel value), and then average that entropy over all possible pixel values (weighted by the probability of getting that pixel value).

Symbolically, we start with some random variable P representing the number of absorbed photons, with some (presumed known) probability distribution $\mathcal{P}(P = n) = P(n)$. In practice, we assume $P(n)$ is a Poissonian distribution. We then have another random variable C that represents the number of extra electrons added by CIC noise, with a Poissonian distribution $\mathcal{P}(C = n) = C(n)$. The combination of these two gives the input value N_0 for the EM stage, and the distribution for that is the convolution of the distributions of P and C, $\mathcal{H}\{N_0\}(n) = \mathcal{P}(P + C = n) = (P(n) * C(n))(n)$. The output of the EM and readout stages is given by our model (for example, Equation (A.56)). Using our model, we can compute a theoretical image histogram for such illumination, $\mathcal{H}\{\text{final}\}(n)$, which gives the probability distribution for the image value I. Now we assume we have an actual image, which yields a particular value of I. Given that, we can compute the probability distribution of the number of absorbed photons using Bayes's Theorem,

[10] Our model does not require that the starting histogram be Poissonian, but since it is in this case, we can generate the histograms using the exact same formulas as for dark frames, and just increase the CIC noise by the number of photons per pixel.

Equation (B.11):

$$\begin{aligned}
\mathcal{P}(P=n \mid I=m) &= \frac{\mathcal{P}(P=n, I=m)}{\mathcal{P}(I=m)} \\
\mathcal{P}(I=m) &= \mathcal{H}\{\text{final}\}(m) \\
\mathcal{P}(P=n, I=m) &= \mathcal{P}(I=m \mid P=n)\mathcal{P}(P=n) \\
\mathcal{P}(I=m \mid P=n) &= \sum_j \mathcal{P}(I=m, N_0=j \mid P=n) \\
&= \sum_j \mathcal{P}(I=m \mid N_0=j, P=n)\mathcal{P}(N_0=j \mid P=n).
\end{aligned}$$

The probability distribution $\mathcal{P}(I=m)$ for a given input N_0 is something we can compute with our model. N_0 actually fully determines that histogram; the $P=n$ condition only matters in determining the probability $\mathcal{P}(N_0=j \mid P=n)$. Since $N_0 = P + C$, and P and C are assumed independent (the CIC and thermal noise should be independent of the signal),

$$\mathcal{P}(N_0=j \mid P=n) = \mathcal{P}(C=j-n \mid P=n) = \mathcal{P}(C=j-n) = C(j-n).$$

Putting all of this together, the probability distribution we are interested in is for the number of absorbed photons given the pixel value in the image:

$$\mathcal{P}(P=n \mid I=m) = \frac{\sum_j \mathcal{P}(I=m \mid N_0=j, P=n)C(j-n)\mathcal{P}(P=n)}{\mathcal{H}\{\text{final}\}(m)}.$$

As a check, if we then average over all possible P values, that is the same as essentially ignoring any actual measurement, so we should recover the original distribution for P:

$$\begin{aligned}
&\sum_m \mathcal{P}(P=n \mid I=m)\mathcal{P}(I=m) \\
&= \sum_m \frac{\sum_j \mathcal{P}(I=m \mid N_0=j, P=n)C(j-n)\mathcal{P}(P=n)}{\mathcal{H}\{\text{final}\}(m)}\mathcal{H}\{\text{final}\}(m) \\
&= \mathcal{P}(P=n)\sum_m \sum_j \mathcal{P}(I=m \mid N_0=j, P=n)C(j-n) \\
&= \mathcal{P}(P=n)\sum_j C(j-n) \sum_m \mathcal{P}(I=m \mid N_0=j, P=n) \\
&= \mathcal{P}(P=n).
\end{aligned}$$

The last step follows because the m sum evaluates to 1 (the probability that I equals some value is 1, no matter what N_0 or P are), and then the j sum evaluates to 1 for the same reason.

Now, to evaluate a camera, we want to know how well-defined our estimate of P is given I, so we want to know some average measure of width of $\mathcal{P}(P = n \mid I = m)$. Thus we pick some entropy-like measure of that probability distribution, and average that over all possible m, weighted by the probability that we will get $I = m$. We tried three different entropy-like measures:

1. The Shannon entropy, defined as $-\sum_n \mathcal{P}(P = n)\ln(\mathcal{P}(P = n))$. In the limit of P being perfectly well-defined ($\mathcal{P}(P = n)$ is 1 for exactly one value of n, and 0 for all other n), this becomes 0, and in the limit of P being evenly spread out over N different values ($\mathcal{P}(P = n) = 1/N$), this becomes $\ln(N)$.

2. The p^2 measure, which we define as $-\ln\big(\sum_n (\mathcal{P}(P = n))^2\big)$. In the limit of well-defined P, this becomes 0, and in the limit of an evenly spread P, this becomes $\ln(N)$, just like the Shannon entropy.

3. The standard deviation. Unlike the other two measures, for a bimodal distribution, this measure gives a value depending on the different between the two different values of P (the other two give $\ln 2$). However, since none of the noise added (including the EM stage) splits the distributions up, we do not see bimodal distributions in this problem, so this, like the other two measures, gives a similar idea of the spread of a distribution.

Using three different measures allows us to check that our results are general results of the spread of the distributions, and not particular to the details of how we are measuring width. If there are large differences in how well a camera can measure an intensity, we expect all three of these measures to give the same rankings, but with possibly different details in cases where the distributions are similar.

We ran simulations of the above entropy measures over a wide range of camera parameters in the vicinity of the parameters we had actually measured in various camera modes, and also produced rankings of the cameras modes we had measured.

These simulations and rankings covered a range of average number of photons absorbed from 0.01 per pixel to 2. We limited the maximum number of photons because we were interested photon-level detection performance, and made several approximations in our formulas assuming the average number of excited electrons before the EM stage was small compared to various other quantities, and we expected our results lose more accuracy for mean values greater than 1.[11] As expected, the various entropy-like measures do indeed give fairly similar results, mostly depending on the CIC levels of the camera. Two such rankings are shown in Figure 5.28 and Figure 5.29. These two rankings were selected because they show how the rankings are similar even for different exposures and entropy measures, as well as some of the differences in ranking we expect from having different exposures, as previously outlined. The low exposure figure (Figure 5.28) shows how the ImagEM values were sorted by camera speed. In agreement with our earlier argument on how cameras would rank for low exposures, this sorting mostly follows the CIC rankings shown in Figure 5.26. For such low exposures, we expect either 0 or 1 photons to hit, and the high amplification pretty much guarantees that these will be distinguishable. The difficulty is in trying to distinguish between actual photons and CIC, and so CIC is the most important factor. The higher exposure figure (Figure 5.29) shows how that sorting starts mixing up, because, at higher exposures, the CIC matters less, and the EM gain relative to readout noise matters more. Larger photon numbers tend to make the CIC less important, and we would expect that, if the input photon number were large enough, eventually the EM gain (relative to the readout noise) would become the dominant parameter for the entropy rankings. Since the EM gain sorts the ImagEM modes in the opposite order by speed (as shown in Figure 5.27), the strong sorting of ImagEM modes by speed in Figure 5.28 is starting to switch has decreased in Figure 5.29. The abnormally high-ranking (by EM gain and readout noise) Evolve:512 point has overtaken the best ImagEM points. We note that one 11 MHz ImagEM point appears to rank quite poorly compared to the other ImagEM points in the high-exposure rankings in Figure 5.29.

[11] We have formulas that do not make this small-excitation assumption in Appendix A. Should the reader wish to do a more thorough computation, everything that is needed should be there.

This point was taken at approximately half the full EM gain. This did not affect the ranking of that point by CIC in Figure 5.26, but did in the ranking by EM gain in Figure 5.27. That point shows us the same transition from CIC being most important to EM gain, as it ranked better in the lower-exposure entropies used in Figure 5.28 then in the higher-exposure entropies used in Figure 5.29.

The anomalous Evolve:512 point ranks as the best point in both Figure 5.28 and Figure 5.29. While it did not have the best CIC value, it was close enough that its much better EM gain ranking was able to overcome that. However, since we suspect the EM gain value was something of a fluke, as described in Section V.11, we would not base any decisions on those point without more data.

Checking Quantum Efficiencies of EMCCDs

The large discrepancy between the best CIC measurements between the ProEM and either the ImagEM, Evolve:512, or iXon seemed suspicious, especially since all the manufacturers use the same EMCCD, and all had camera modes where they optimized for CIC. In discussing possible causes for the discrepancy, it was brought up that the Hamamatsu software we were using with the ImagEM had an option for processing the image to lessen the effects of CIC in an image. We tracked that option down, and verified that we were not using it. Still, the question remained, could there be some image processing that artificially reduced the CIC?

To test this idea, we used our histogram fits to measure the quantum efficiency of an EMCCD. As this testing is more complicated than just taking dark frames, we only have data for some of the demo cameras that we actually had in the lab, restricting this discussion to ProEM and ImagEM cameras.

Our model assumes that the EMCCD histogram, before entering the EMCCD, is a Poissonian distribution resulting from CIC (and thermal noise, as appropriate). However, a highly attenuated laser beam (near the single-photon limit) should also produce Poissonian statistics. If we illuminate the EMCCD with a very faint pulse of laser light, the distribution of the electrons in the cells as they enter the EM stage will be from the sum of the absorbed photons with the CIC. Since an absorbed

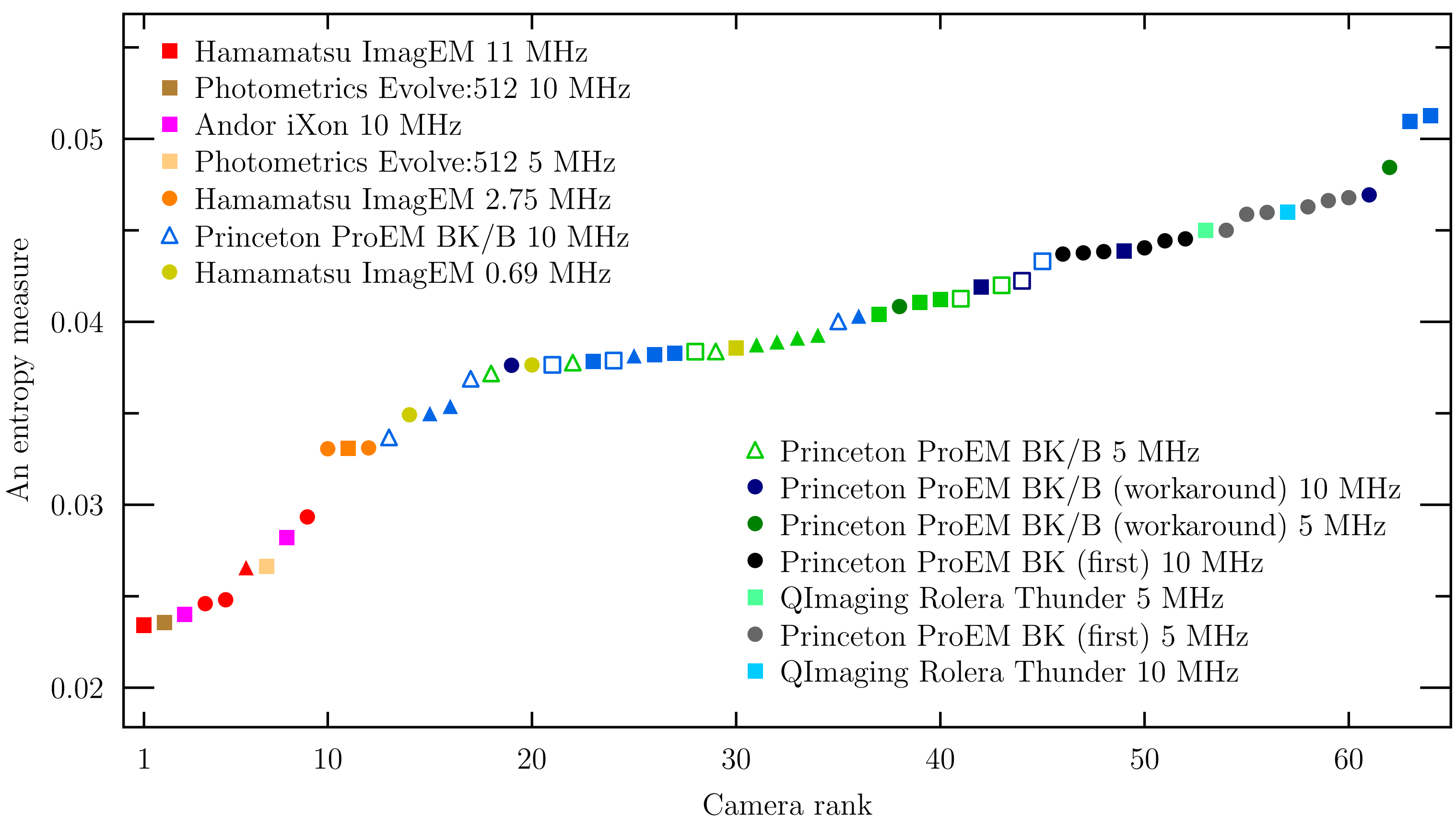


Figure 5.28. The expected Shannon entropy at 0.01 photons/pixel for various cameras in various imaging modes, ranked from lowest to highest. We computed the expected entropies assuming an average illumination of 0.01 photons/pixel/exposure. The color scheme and shape scheme is described in the text under Figure 5.26.

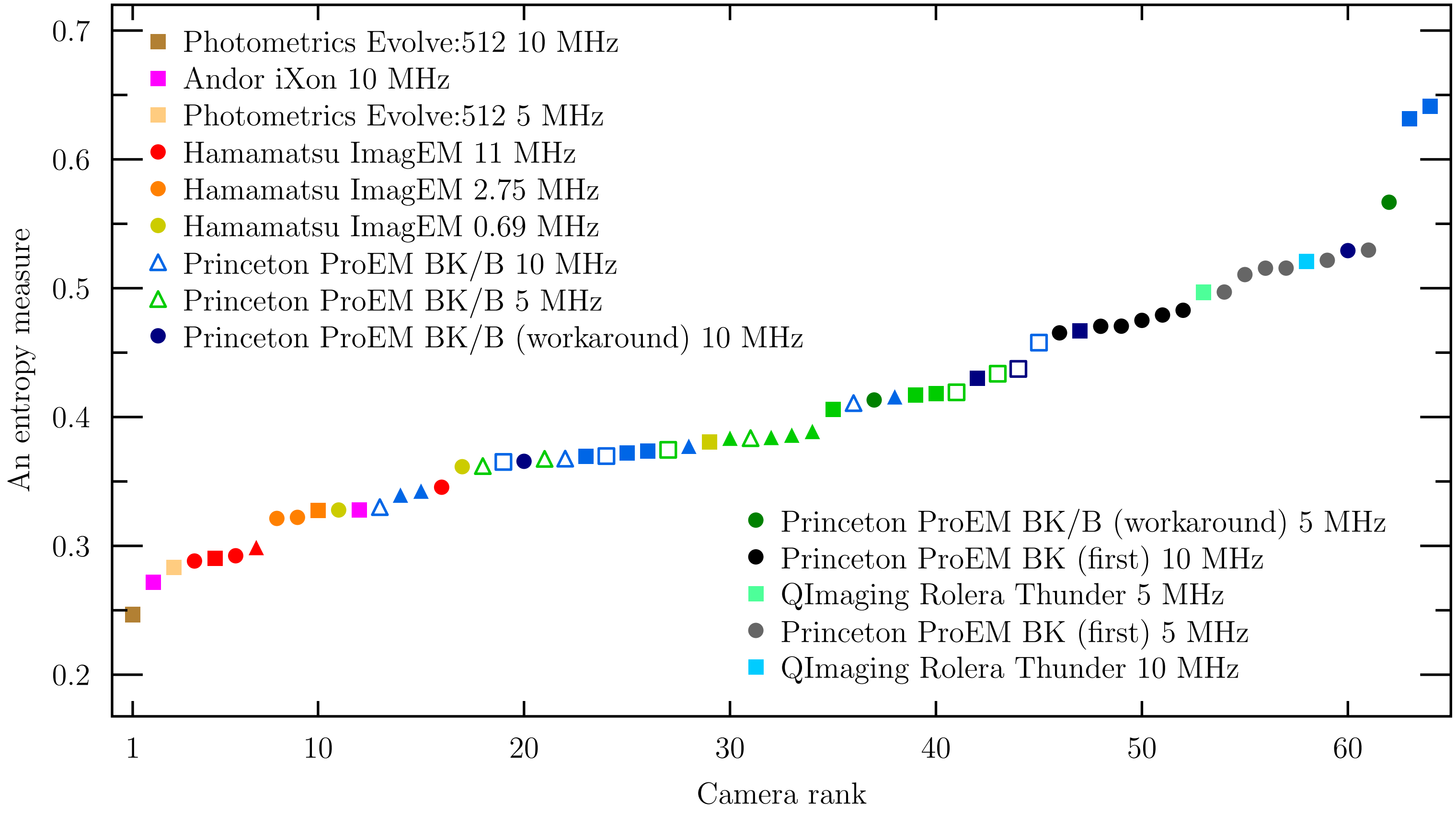


Figure 5.29. The expected p^2 entropy at 0.5 photons/pixel for various cameras in various imaging modes, ranked from lowest to highest. We computed the expected entropies assuming an average illumination of 0.5 photons/pixel/exposure. The color scheme and shape scheme is described in the text under Figure 5.26.

photon excites a single electron, and random absorption of a Poissonian source is also Poissonian,[12] and the sum of two Poissonian variables is also a Poissonian variable (as shown in Section B.4), the pre-EM histograms should still be Poissonian, but with the mean being the sum of the CIC and the light signal times the fraction actually absorbed (the quantum efficiency) and the CIC mean. Thus, if we use our dark-frame histogram model to measure CIC from a series of images where there was a constant low-intensity feature, the CIC value will actually be the sum of the CIC and the mean value of absorbed photons per pixel. Thus, our measured "CIC" value should increase linearly with the number of photons actually incident, with the slope being te quantum efficiency.

What is more, if there was some processing going on that made an image look like there were less CIC[13] it should also decrease the effective CIC from the incident light, and we should see a smaller quantum efficiency.[14]

We tried this experiment. We started with a relatively weak laser beam (around 20 μW) with a $1/e^2$ intensity radius of between 2 mm and 3 mm, and passed it through a series of neutral density filters to decrease the power to about 26 pW.[15] This corresponds to approximately 1×10^8 photons/second, with a peak intensity on the order of 2×10^4 photons/pixel/second, given that the pixel on the EMCCDs we are dealing with are 16 μm on a side. Thus, by varying the pulse length from tens of microseconds to a few milliseconds, I can cover a range from well under a photon per pixel to several photons per pixel.

Before the filters, the beam passed thorough an iris with a diameter of about

[12] After all, it satisfies the requirements to be Poissonian, namely that there are a series of random, independent events with a fixed probability of occurring in any given time interval.

[13] . . . and amazingly not change the shape of the histogram . . .

[14] If there were some processing, but somehow it only affected true CIC and not actual light signals, and it affected the CIC in such a way that the histograms look just as if CIC were just decreased, then that would not be a problem.

[15] We measured the power reduction of the filters at much higher beam powers, so that the transmitted power was large enough to be well above the threshold of our power meter. We verified that the transmittance was the same for a range of powers, and assumed that it would also be the same for low incident powers, where the transmitted beam was too weak to measure with our power meter.

3.9 mm, which is small enough so that only the center, roughly constant-intensity part of the Gaussian beam passed through. The whole assembly was then pressed up very close to the window over the EMCCD on the camera, and the edges were sealed so that no light could get into the camera save through the neutral density filters. There were no lenses on the camera; light passed directly through the iris and neutral density filters onto the EMCCD, giving us a roughly circular beam profile of constant intensity. Figure 5.21 shows images of these beams, without the irises. We used these images to determine the size of the beam, and to determine what region we could use as a region of roughly constant intensity.

Next, we imaged the beam profile at maximum EM gain, with 50 ms exposures, both with and without the iris, and used Gaussian fits with circular masks to estimate the beam waists and iris size. These allowed us to infer which part of the beam was passing through the iris, and verify that the selected part of the beam was roughly constant intensity. Combining that with the known intensity of the beam, we can convert that to the average number of photons hitting each pixel per second.

The beam we used was the same one we used for absorption imaging of the atoms, allowing us to make use of the pulse stability circuit in Section III.10. The beam is shuttered with an AOM and passes into a fiber. The output of the fiber passes through an uncoated 2° wedge and hits the photodetector of the pulse stabilizer circuit, while the reflection off the first surface was redirected through the iris and neutral density filters onto the EMCCD. While the range of pulse fluxes we wanted was too large for the pulse stabilizer circuit, we were able to monitor the photodiode output with an oscilloscope to measure the total flux of each pulse, which we could then convert to an expected number of photons incident on each illuminated pixel of the EMCCD.

Once this setup was in place, the rest was conceptually simple. With the room lights off, we took images with a range of laser pulse lengths (verifying that all of the pulses had very close to the same total flux). We then isolated two regions of the EMCCD, one that was entirely in a nearly constant-intensity illuminated region, and one that was outside of the illumination region, far from any fringes. For each region and each pulse length, we computed aggregate histograms in the same way we did for dark frames, performed the dark-frame histogram fit, and looked for correlations of the fit parameters with the expected number of incident photons per pixel.

The data for the ImagEM EMCCD is shown in Figure 5.30, where we used variations in the measured pulse flux to estimate the error in photons per pulse (which is too small to see in the figure), and left off any error estimate in CIC (but 10% is probably a good starting point). The dark region remained dark, while the light region showed the correct CIC value with no incident light, and showed an increase in "CIC" of about 0.73 e^-/incident photon. The quoted quantum efficiency of the EMCCD at 780 nm is about 75% [101, 102]. Either we got very lucky and possible errors in laser power, beam size, pulse flux, conversions to photons per pixel, curve fits, and any errors introduced by some filtering of the camera all combined in such a way as to mutually cancel and reproduce the quoted value for the EMCCD, or, if we did everything correctly, we have found that any filtering done by the camera does not affect our ability to count photons absorbed by the EMCCD.

As an aside, we also checked the other fit parameters to see if they changed with incoming light. Readout noise increased in a fairly linear fashion by under half a percent as the input light increased to about 1 photon. Above that, the increase rate seemed to grow a lot. EM gain was difficult to measure well for these. EM leakage also seemed to increase as flux increased (by about 70% of the flux value), but dropped mostly to zero when for larger input fluxes. The illuminated region did not have as many pixels as a full frame, and, since we took the same number of images, the total number of pixels we used for the histogram was much smaller than for the dark frames. Since relatively few pixels form the CIC tail in the darker histograms, this made for a very noisy tail, making it hard to determine the EM gain well. The EM leakage growth may have been a real effect, but the zero values were probably not. The biggest effect of EM leakage on the histograms is to round out the elbow between the CIC tail and the main Gaussian. At higher input values, the Gaussian peak is not as dominant a feature, since fewer pixels enter the EM stage with no electrons, and there is more spread among higher values of input electrons.

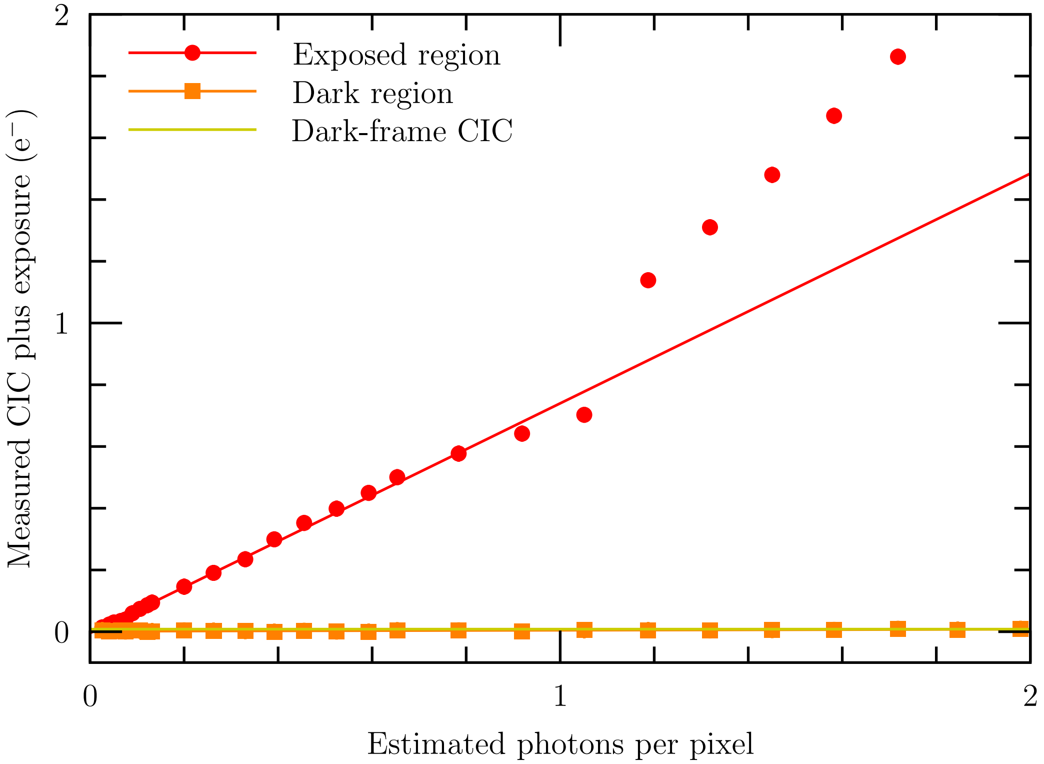


Figure 5.30. A measurement of EMCCD quantum efficiency. The two sets of points show the CIC from the curve-fits as functions of the estimated photons per pixel, with the lines of best fit. The horizontal error bars are too small to see on this plot, while the vertical error bars are left off, but should be of a similar magnitude as described in the caption of Figure 5.26. One curve is from the non-illuminated region of the EMCCD, and shows essentially no dependence on the incident light level, which tells us that dark regions of the EMCCD are not affected by other regions being illuminated (at low intensities). The second curve is from the illuminated region, and shows the CIC increasing as a function of the incident light. The y-intercept of the line agrees with both the dark-region CIC and a full dark-frame CIC measurement for the camera. The slope of the curve tells us how many electrons, on average, were excited by the light for each incident photon. That slope is about 0.73 e^-/photon, indicating that the quantum efficiency of the EMCCD is about 73%. We note that the CIC dependence on input flux is no longer linear past about 1 photon. We talk more about this in the main text, but this is probably either an effect of some approximations we made in deriving our dark-frame model where we assumed the CIC would be under 1 electron per frame, or a result of variations in the intensity of the illuminated region.

With no sharp elbow for EM gain to round out, there is not much to determine EM leakage values that cannot be mostly fit by varying the CIC. Also, as we explain shortly, there are other reasons not to trust the larger input flux points. As for the lower values, we think it quite possible that running more current through the EM stage could cause it to leak a little more, possibly through a mild heating. It is also possible that the curve-fits have some trouble differentiating between EM leakage and pre-EM CIC, and that part of the apparent CIC increase may be treated as EM leakage. In any case, the leakage values alter the CIC.

The initial increase in readout noise may be real. For example, the hysteresis mentioned in Figure 5.35 could explain it. If the readout amplifier is jumping around more due to more input signal, the hysteresis could smear these values together, which is effectively an increased readout noise. The larger increase at larger input fluxes, however, we think is not real. Our explanation also explains why we do not trust the curve-fits for the higher fluxes, even though the model seemed to fit decently. First, we note that the illuminated region is not illuminated in a perfectly constant manner. By taking a histogram over the entire region with a range of mean exposure values, we are essentially averaging histograms with different CIC values together. At lower mean values (smaller exposures), the post-EM histograms look like exponential tails, and averaging several tails with different areas under them (but the same length scale) results in a correct CIC value. For larger mean values (larger exposures), the gap between the maximum and minimum exposure is larger. Also, the Poissonian distributions look more like a Gaussian than an exponential, and after the EM stage starts looking more like a peaked curve than an exponential tail. Averaging over several mean values, each with a different peak, effectively broadens that peak. The result of such an averaging would look like an increase in readout noise, and may affect other parameters of the model as well. A second thing to note is that, to aid in curve-fitting, we made a few extra approximations that assume the mean CIC value (here, that includes electrons from the input photons) is about 1 or less. At higher input photon numbers, these approximations become less valid, and the curve-fitting routines may be altering the model parameters incorrectly to compensate, apparently with some success. Thus, we think we can explain the extra increase in readout noise for larger input flux values, and why we do not feel we can trust the curve-fits for these inputs. In particular, this probably explains why the CIC values for larger fluxes in Figure 5.30 no longer follow a linear relation.

We also did the same analysis with the one of the ProEM demo cameras we had. However, that camera had strong shading problems (much larger than the other cameras we had), along with a few other odd artifacts that showed up in this measurement, including some odd horizontal stripes that did not show up in the dark-frame measurements.[16] These effects led to rather poor curve-fits, and the plot corresponding to Figure 5.30 is a lot noisier and does not appear to be as linear. As we mostly wanted to check that the curve-fit method could measure quantum efficiency, and verify that the ImagEM was not artificially enhancing their CIC value (in such a way that would prevent us from counting photons accurately), we did not pursue the same measurement on newer ProEM cameras.

EMCCD Anomalies

We will close this chapter with a brief discussion of some of the anomalies we found while learning about and modeling EMCCDs, and trying to compare cameras in a meaningful, quantitative way. We present these problems in the hope of giving somebody else an idea of things to watch out for when evaluating an EMCCD camera.

Figure 5.13 shows an obviously clipped histogram from a Hamamatsu ImagEM camera. While it makes the dark frames look nice, as seen in Figure 5.12, and is really only a problem in very dark frames, it makes it difficult to determine the readout noise of the camera, since the main Gaussian peak is the narrowest pre-readout-stage feature we could hope for (even illumination would probably have

[16] These stripes may be related to the first-of-sequence problem mentioned in Section V.8 and Section V.11, as we did not know about the problem at the time and so were not correcting for it. This camera compared poorly even to the later ProEM demo cameras we had, including others where we were not correcting for the same problem, so it may also have been a problem with this particular camera.

more spread). It would also make it slightly harder to differentiate between 0 electrons and 1 electrons, as many 0-counts, and a very small fraction of 1-counts are combined into a clipped-off spike. However, since such a small fraction of the tail is lost, the clipped peak is almost entirely 0-count, so we do not feel this is much of a problem. Furthermore, we found it was only a problem with the higher gain settings of the readout stage, at higher EM gains. At high EM gains, there is not much incentive to use high readout gain settings, since those do not amplify over the readout noise much (both signal and noise go up almost proportionally), especially when compared to the effect of the EM gain. The main reason, we think, to use the higher readout gain settings is a low EM gains, where this clipping does not seem to occur. At low (or no) EM gains, a low readout gain can result in a small signal being lost in the non-zero resolution of a single ADU, although this effect is unlikely since even then the readout noise is larger than an ADU. At higher EM gains, the amplification of the EM stage is so large compared to a single ADU that there is no reason we know of to use the readout stage to amplify even more.

Another problem we mentioned earlier is a change in readout noise. As described in Section V.8, we found our measured value for the readout noise in the Princeton Instruments ProEM cameras increased dramatically as the EM gain was increased. Since, if one really cares about reducing readout noise, one will probably use a higher EM gain setting, it seems that the most useful readout noise is one at higher EM gains (except for the modes with no EM gain at all). Since this is not the number quoted, we do recommend anyone considering using an EMCCD make their own readout noise measurement, in case the readout noise for the particular mode and camera they might use depends on the EM gain. We also found this problem in the iXon, Evolve:512, and Rolera cameras. The Hamamatsu ImagEM did not seem to have this problem.

Yet another odd datum is the Photometrics Evolve:512 10 MHz point in Figure 5.27. Here, we see the EM gain divided by the readout noise for the 10 MHz Evolve:512 data point is much larger than any other point, including the 5 MHz data point for the same camera. This seems counterintuitive, because both points should have the same EM gain of 1000, and the slower readout rate should a lower readout noise. Thus, the slower readout rate should have the higher ratio. As we did not take these images ourselves, nor did we ever have access to the camera that took them, we are restricted to the data we were given to guess what is happening here. Fortunately, we do have a few extra data sets. One of these sets was supposedly taken in the exact same mode, but with the EM gain turned down to near-unity. This mode should have the exact same readout amplification, so we can compare the readout noise in ADU for these two data sets. In ADU, the readout noise increases as the EM gain is turned up from 1 to 1000, just like with so many of the other cameras. If we assume that, like the other cameras, the quoted typical readout noise is accurate for the low EM gain, then that allows us to get an ADU-to-e$^-$ conversion for this mode. We can use that to infer the actual readout noise and EM gain for the 10 MHz data point. We find that the 10 MHz data point appears to have a readout noise of about 65 e$^-$ (about 1.4 times larger than the data set with low EM gain), and an EM gain of slightly over 2100. The readout noise increase seems in line with what we have seen for other cameras, which suggests it really is the EM gain that is about 2.1 times larger than the maximum specification. Even if we assume that the readout gain increased as the EM gain increased, by a factor of 1.4 (so that the readout noise actually stayed constant), then the EM gain would still be about 1500, which is still quite large. The ADU-per-e$^-$ we get from this method is about the same as we get if we do the same with our two 5 MHz data sets, one with EM gain near unity, and the other with an EM gain of 1000. If we assume the 5 MHz EM gain is actually 1000, then the near-unity EM gain readout noise is about 80% larger than the quoted value. If we make a similar assumption for the 10 MHz data, then we need to decrease our ADU-per-e$^-$ value by about 1.8, which would make the EM gain closer to 1200, which is a bit more realistic. Two conclusions seem highly likely. First, the maximum EM gain for the 10 MHz mode *for this particular camera* is more than for the 5 MHz mode, and probably more than the quoted maximum of 1000. Second, either the readout noise or the EM gain for both modes is pretty far from the specifications.

It appears at least one thing, and probably several, are rather unpredictable with the Evolve:512 that was used to take the images we analyzed. Our request for comments on this effect from the representatives was not returned, so we are left with our own speculations. On several of the cameras, we have found the specifications often underquote the readout noises, so we suspect that effect may be part of what is happening here, but we must also conclude that the EM gain was not very well calibrated on this camera. The readout noises are probably larger than the specification, and the EM gain is apparently at least capable of being larger than the quoted maximum. That poses an interesting question: Why, if the EM gain could be larger, do the camera manufacturers not run it at the maximum value? We have a few speculations, but they are just our speculations, and nothing more. First of all, new cameras tend to have higher EM gains. As they are used, the EM gain drops, and eventually, the EM gain needs to be recalibrated. This happens more rapidly at the beginning of the camera's life than towards the end.[17] It also happens more rapidly when the camera is run at a higher EM gain. Thus, we assume that most cameras are actually run at less than their true maximum EM gain. This way, as the camera ages, the EM gain does not decrease too rapidly. It also allows for the EM gain to be re-calibrated. If the camera were running at the true maximum EM gain, and the true EM gain dropped, no amount of calibration could restore that maximum EM gain. If, however, the EM gain were never run at the true maximum, then, as the EM gain dropped a little, the camera could be re-calibrated, since there would be some freedom to still increase the EM gain a little bit more. We thus speculate that EMCCD cameras are typically not run at their true quoted maximum EM gain, because the camera manufacturers prefer to have a camera where the quoted maximum EM gain can be maintained over the lifetime of the camera, rather than a camera where the maximum EM gain, while higher, does not last.

Perhaps the most startling problem we encountered is what we have been calling the first-of-sequence problem. The problem itself is covered in Section V.8, so we will focus mostly on how we discovered it here.

[17] For this reason, Hamamatsu "ages" new EMCCD cameras by taking many exposures with the EM gain turned on [38]. This gets the cameras past the stage where the EM gain degrades rapidly.

Our first step towards finding this problem was when we tried to reconcile why the ProEM EM gain (relative to readout noise) did not easily outperform the ImagEM in our EM gain rankings. As we mentioned when describing Figure 5.27, the ImagEM 11 MHz modes should have had relative EM gains of about 12, which they did, while the ProEM 10 MHz modes should have had a relative EM gain of closer to 20, which they did not. Our fits gave us the EM gain in ADU, and the ProEM cameras came with specification sheets that quoted ADU-to-electron conversions specific to each camera, so we were able to come up with actual EM gains from the fits. We found the maximum gain was quite close to the quoted value of 1000, and so we figured the problem was in the readout noise. We then questioned some contacts at Princeton Instruments about this, and learned their methods for computing readout noise. As discussed in Section V.8, we tried to duplicate their readout noise measurements.

Since the method Princeton Instruments uses to measure readout noise does not work for large EM gains, we used low-EM gain and no-EM gain modes to compare with our curve-fits. Since these modes do not produce CIC tails (with a gain of 1, the tail is buried in the readout noise), our fits fail to produces useful values for EM gain and CIC noise in these fits, but still fit a Gaussian to get decent values for readout noise, so we can directly compare these. It was from these comparisons that we noticed both that the readout noise for these modes was less than at the high EM gain modes, as measured by our curve-fitting method. We initially had trouble getting our values to agree with the values quoted by Princeton Instruments, until we plotted many such measurements, such as in Figure 5.25. Then, we realized that the quoted value seemed to be the minimum of a rather large range of possible readout noises, depending on exactly which two images we subtracted.

This lead to us contacting Princeton Instruments again, mentioning the large spread in readout noises (in particular, how much larger that spread was than the for the ImagEM), and how their quoted values seemed to be the minimum rather than the average value. We sent them some of our data so they could confirm our results, and from that, they were able to diagnose our problem.

When we took image sets, we used software options that took the images with slight delays (typically one second) between the images, so any odd correlations between images would have a chance to disappear before we took the next image. This seemed reasonable, since in our experiments, there is typically a long wait (on the order of tens of seconds—"long" is relative to the maximum frame rate of the camera). We are mildly amused, therefore, that in doing so we caused the opposite problem. The software provided with the ProEM implements the delay by starting and stopping the camera, so that each image is effectively the first in a sequence. On the ProEM, the first image of a sequence is apparently taken before the readout electronics have fully settled, resulting in excessive readout noise and a larger offset between the two readout amplifiers (we will discuss these later in this section), and stronger shading. The effect varies with each first image, and so results in highly varying results, as seen in Figure 5.25.

Our Princeton Instruments contacts offered two solutions. The first was to set up the cameras to take all the images from a single sequence. The second was to alter the settings that determined how the CCD was cleaned while waiting to take an exposure. The first solution worked, in that the first image had the problems we had been seeing, but the rest seemed to be fine. The second solution did not work. We did not check if the first-of-sequence problem was solved by the second method, as that method caused the images to appear as though the center of the image was somehow exposed to a little bit of light, which was not acceptable to us. Naturally, we took images at full EM gains with the original clean settings, and there was no light evident. Furthermore, as we were already in a dark room with the lens cap on the camera, we put the camera in a dark box in the dark room, and the exposure effect was still there, unchanged, whenever the cleaning settings were altered. As a result of this, we did not use the second solution, and only looked at results from the first, with the results we have been using in this chapter. Typically, we would throw out the first image, but in some of the histograms (like Figure 5.14), we left it in to show the effects of the problem.

We did not investigate enough to determine whether this would have been a major problem for us, as even with the fix, the improved readout noise of the ProEM was not enough to compensate for the much-improved CIC of the ImagEM.

However, we admit that there is a possibility that the first-of-sequence problem would not occur in our normal experiments. It is possible that the first image was only bad if the exposure happened immediately after the start of the sequence. If that were the case, if we were to start an image sequence to take an exposure upon a trigger at the start of each repetition of an experiment, but the actual trigger and exposure happened several seconds later, the problem would not occur. Since that is the way most of our experiments work, the problem might not manifest itself in our normal experiments. If that is not the case, we could probably work around it by always taking pairs of triggered images, with a fake trigger to take a junk image at the beginning of the experiment repetition, and a second trigger to take the data image later, and maybe a third dark image afterwards. It is likely that we could have found a suitable workaround. No such workaround seemed necessary for the ImagEM, as it did not seem to have this problem. Also, the correlations images are scaled badly, and are all black.

The first ProEM demo camera we acquired seemed to have abnormally large shading compared to the other ProEM demo cameras. The magnitude of this shading issue can be seen in Figure 5.31, which shows a sample dark frame, and how the shading the very noticeable over the readout noise. Since this shading did not subtract out, as it was not constant from image to image, this resulted in our curve-fits reporting large readout noises for this camera, and making it difficult to fit our model to the image histograms in certain cases. This camera also had large CIC noise, which accounts for the exceptionally bad rankings of this camera in Figure 5.27 and Figure 5.26. While these effects were quite a bit worse than we saw in later demo cameras, we note that we did not know about this first-in-sequence problem while we had possession of that camera. It is likely that had we corrected for that first-in-sequence problem, a large part of the extra noise would have disappeared, and it is possible that the camera would have then performed closer to the level of the later demo cameras.

Another problem we noticed with the one of the demo ProEM cameras was when we attempted to measure the quantum efficiency of the camera, as described in Section V.10. As we did not repeat this measurement with other demo cameras,

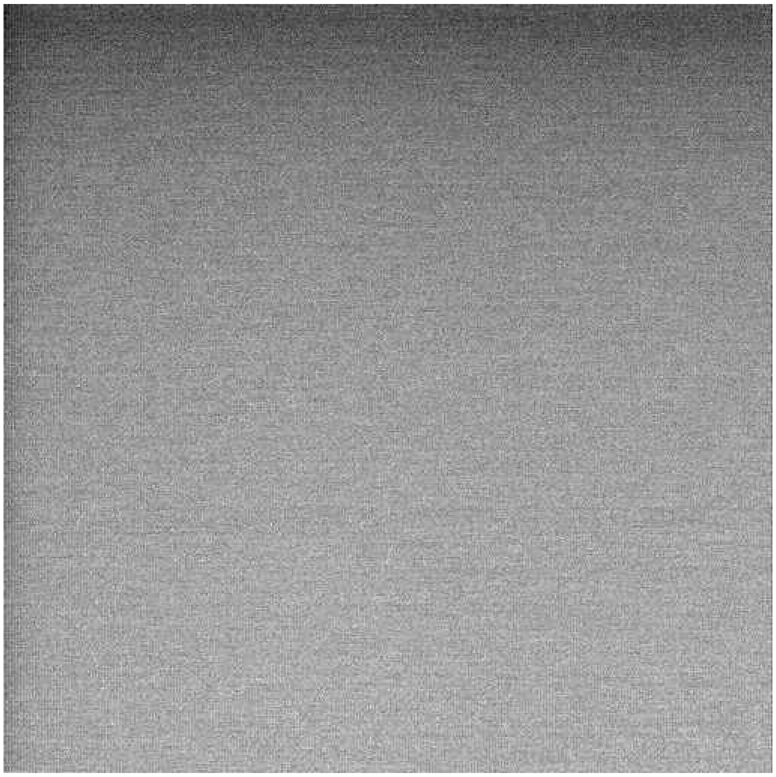


Figure 5.31. A dark frame with strong shading. This dark frame shows the readout noise and strong shading on the first ProEM camera we had as a demo. The grayscale has been adjusted to the scale of the fluctuations which, while small compared to the dynamic range of the camera of the signal from a fairly faint source of light, are large compared with the readout noise. This image was taken with a low EM gain, as otherwise the multiplied CIC alters the scale, and so should be compared with the much less shading in the no-EM ImagEM image in Figure 5.8.

we do not know whether the problem was specific to this camera, related to the unusually large noise of this camera, or if it could have been fixed by correcting for the first-of-sequence problem. We do not know the cause, and so will simply point it out as something to look for if one is evaluating an EMCCD camera.

We only noticed the problem when we were taking cropped images of the very faint laser pulses which we used to determine quantum efficiency. Figure 5.32 shows an example of this. We see what is essentially a dark frame, with a roughly circular region of very faint, nearly constant intensity illumination from the clipped laser pulse. Superimposed on the image is what appears to be a region where the readout stage had a temporary, abrupt shift in offset, resulting in a horizontal rectangular region. These regions did not occur in every image, nor in the same place when they did occur. We also did not see this effect in uncropped dark frames. In theory, we could have detected these shifts and corrected for them, but, as mentioned in Section V.10, we were mostly interested in the results for the ImagEM, and did not consider it worth our time. Thus, images such as these messed up the curve-fits, and prevented us from getting a good measure of the quantum efficiency of the ProEM cameras. Even with these problems, however, we still got a very rough measurement of about 96%, with enough noise that we still consider this consistent with the expected value of around 70–80%.

We will end this section with what we consider to be the most interesting anomaly. Rather than being a problem, this is actually more of a useful diagnostic, which actually revealed something about the ProEM camera that we had not originally known. When inspecting some ProEM dark frames using high-contrast colors, we noticed that the pixel-columns seemed to alternate between a brighter value and a darker value. Figure 5.33 includes a zoom of such an image, but not one of the more obvious ones. It was as though the readout electronics alternated between two slightly different readout offsets as the pixels were read out. The effect did not seem to be in all images. To investigate, we made some correlation images. We first started with an image, either a raw dark frame, an average of many dark frames, or the subtraction of two dark frames. We then subtracted out the mean value of that image, and computed the correlation image. Let $I(i, j)$ be the value of the mean-subtracted image at row i and column j, and let $C(n, m)$ be the value of the correlation image at row n and column m. Then, $C(n, m)$ is simply $I(i, j)I(i + n, j + m)$ averaged over every pixel of the mean-subtracted image. Any time a coordinate exceed the dimensions of the image, it is assumed to wrap back around to the other side.[18] Note that $C(0, 0)$ is just the variance of the image. In

[18] Note that this computation, as described, is $\mathcal{O}\big(N^4\big)$ for an N-by-N image ($\mathcal{O}\big(N^2\big)$ for a single pixel of the correlation image, and another factor of N^2 to compute every pixel of the correlation image, if you chose to compute the entire correlation image). The correlation image is just a convolution of the mean-subtracted image with itself inverted through the origin, so we can apply

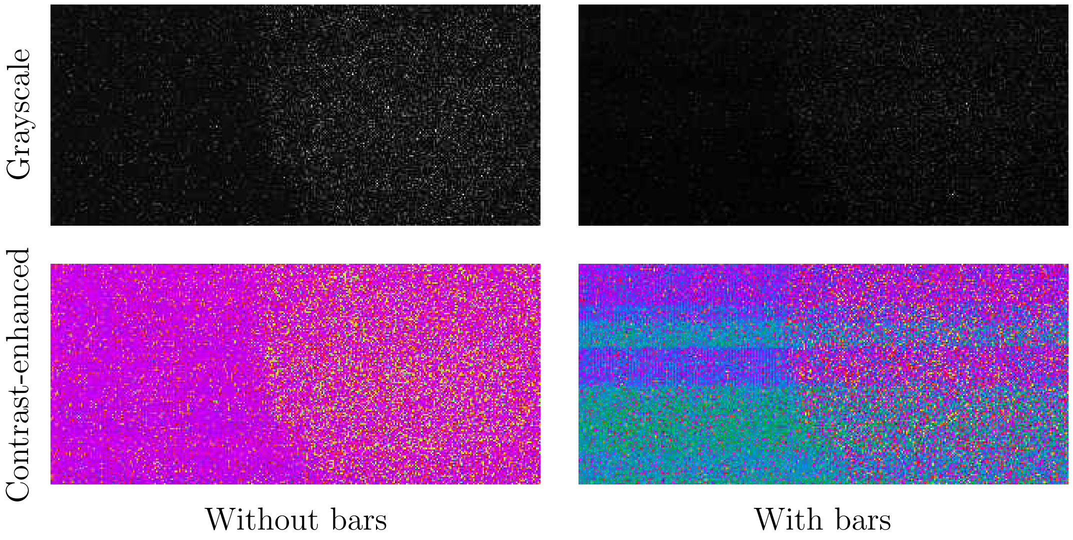


Figure 5.32. A dark frame with horizontal bars. This shows two images taken while we were trying to measure the quantum efficiency of a ProEM camera. The left side shows a normal image, both in grayscale and using a high-contrast color scheme. The right side shows an image with odd horizontal bars across it, in both grayscale and using a high-contrast color scheme. The bars do not really show up in the grayscale image, but the color scheme really brings them out. The bars appear to be regions where the background offset had suddenly shifted, and not the same between the two readout amplifiers, resulting in different levels of the column effect shown in Figure 5.33. This effect happened many times during the quantum efficiency measurement.

order to help comparisons between images, we normalize the correlation image to this variance, so that $C(0,0)$ is 1. In the absence of EM gain, the standard deviation of the image (in the absence of any shading) is the readout noise. In the presence of EM gain, with non-zero CIC, the standard deviation is still on the order of the readout noise, so, in these normalized images, the square root of the magnitude of

the Convolution Theorem. Here, that means the Fourier Transform of the correlation image is the squared-magnitude of the Fourier Transform of the mean-subtracted image. Using a Fast Fourier Transform algorithm, we can compute the Fourier Transform (and the inverse transform) of the main image as a $\mathcal{O}(N^2 \ln(N))$ operation, while the squared-magnitude is a $\mathcal{O}(N^2)$ operation. Thus, by working in Fourier space, we can reduce computing the correlation image to a $\mathcal{O}(N^2 \ln(N))$ operation. If we were to choose to compute the correlation image for the entire image Since $N = 512$ for us, if we were to compute a large fraction of the correlation image, this trick speeds up the computation by well over a factor of a thousand.

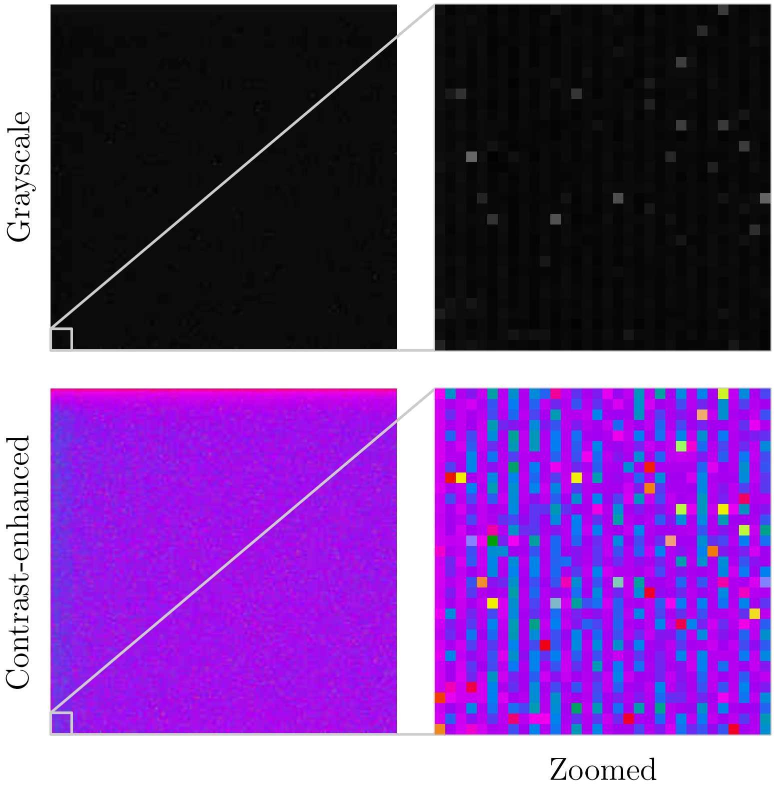


Figure 5.33. A dark frame with vertical columns, from a ProEM camera. The left images show the entire dark frame, while the right images show the same image, zoomed in to the lower-left corner. The upper images are in grayscale, while the lower images are the same two images using a high-contrast color scheme. Note the slight shading in the entire image, and the column effect which dominates the readout noise along the left edge. The column effect can be seen throughout the image, but is strongest along the left edge in this image.

any pixel gives a rough comparison of how strong the (anti-)correlation is compared to the readout noise.

Figure 5.34 shows the correlation image associated with the dark frame in Figure 5.33. In the original image, columns of the dark frame alternate being slightly above or below the mean value of the image. Thus, in the mean-subtracted image, they alternate between slightly negative and slightly positive. Thanks to the readout noise, this is hard to notice in the original image. In the correlation image, since we average over the whole image, the readout noise is suppressed, but since the anti-correlation between columns is constant, it becomes the dominant feature of the correlation image, showing up as alternating vertical stripes in the image.

The column effect shown in Figure 5.34 is very weak; note that the square root of the magnitude is about 1% of the standard deviation from the readout noise. Thus, it is not a significant contribution to readout noise. In fact, it turns out it is an artifact of something that actually *reduces* the readout noise. Through correspondence with contacts at Princeton Instruments, we learned that the ProEM actually has two readout amplifiers, allowing them to read two pixels at a time [40, 102]. Thus, for example, with a readout rate of 10 MHz, they actually run at 5 MHz, clocking horizontally twice with each time step, and reading two values simultaneously, so that the average rate of reading is 10 MHz. In general, the longer a readout has to average a measurement, the lower the noise tends to be (note that all the cameras specify less readout noise at slower readout rates). Since with two readout amplifiers, they get to get to spend twice as much time reading each pixel as a single amplifier would at the same readout rate, having two readout amplifiers can reduce readout noise. The column effect shown so well in Figure 5.34 is a very sensitive measurement of a very slight relative offset between the two readout amplifiers.

We learned that the effect was much more pronounced in first-of-sequence images, as the relative offset is apparently one of the things that needs to settle. In fact, the offset is large enough that it is actually comparable to the readout noise in the first images. That, combined with the fact that the offsets drifts from first image to first image (and thus does not subtract out), is why this effect was strong enough for us to notice while looking at the dark frames, and why it seemed to come and go.

There is one more effect of interest in Figure 5.34. Note the streak of slightly positive correlations that runs horizontally through the center (and, to a much lesser extent, vertically). We believe this streak represents a slight hysteresis in the readout amplifiers. This indicates that if a readout amplifier reads a certain value, the next pixel is slightly biased to the same value. This could be due to any number of effects (like a small amount of charge being left over from a readout), and seems

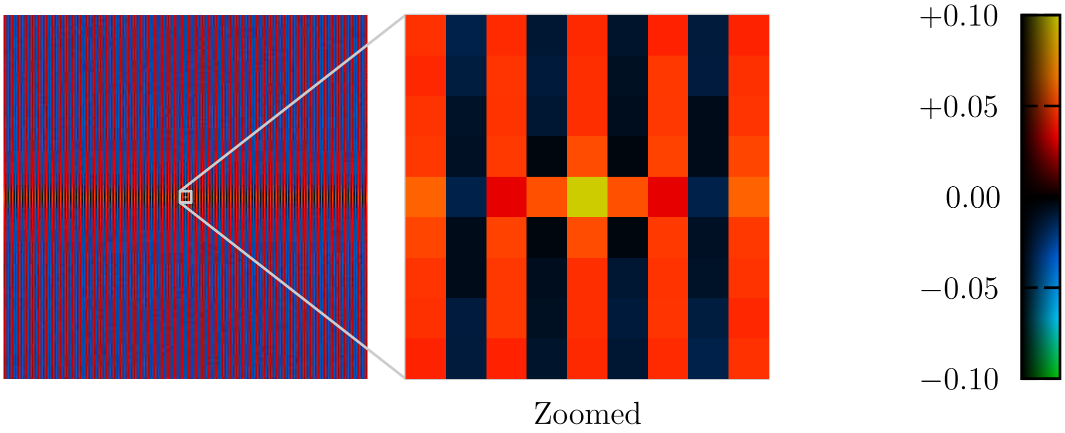


Figure 5.34. Correlations in a ProEM dark frame. This correlation image was computed from the dark frame shown in Figure 5.33, and is normalized to the variance in the image, which is approximately the variance associated with the readout noise. The center pixel is the average correlation of every pixel with itself, which is the variance and hence normalized to 1, but we scale our color scheme to show the rest of the correlations. The pixel m columns over and n columns down from the center is the average correlation of every pixel with the pixel m columns over and n columns down from itself (and so this image is symmetric about the origin). We show both a large-scale correlation image, which shows correlations over about a quarter of the image, and a zoomed-in correlation image, showing individual columns. The correlation image greatly amplifies the column effect shown in Figure 5.33, as the readout noise variations average out, while, since *every* column is slightly anti-correlated with the adjacent columns, that effect remains. On the large scale, we can see wide, not-very-tall region through the origin shows a slightly more positive correlation than the rest of the image. We interpret this as showing slight hysteresis in the readout electronics. In the zoomed image, the positive columns have a larger absolute value than the negative columns. Ordinarily, they have the same absolute magnitude. However, the strongest shading in the image, which can be seen in Figure 5.33, is mostly a slightly dimmer band across the top of the image. This causes every row to be somewhat partially correlated with nearby rows, adding a slight positive offset to a band of rows through the center of the correlation image. We can see this in the larger correlation image, and this offset causes the asymmetric columns in the zoomed image.

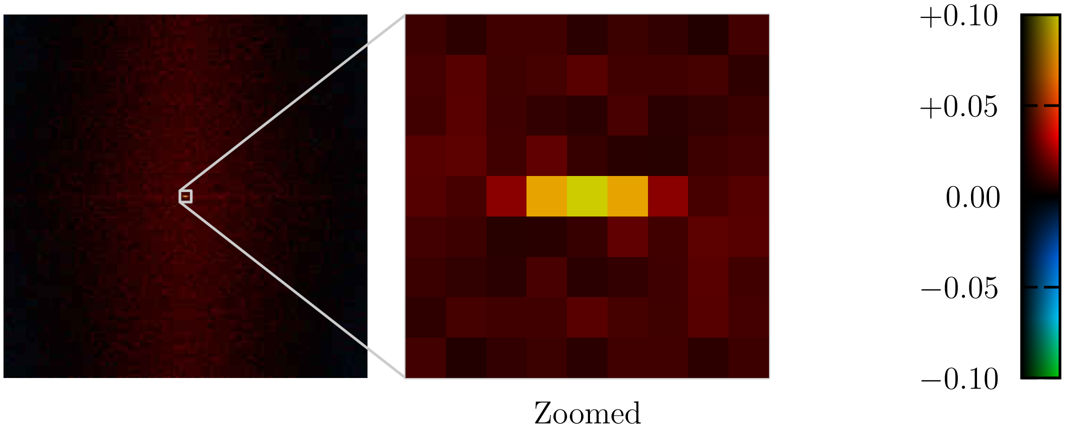


Figure 5.35. Correlations in an ImagEM dark frame. This correlation image was computed from the dark frame shown in Figure 5.9, computed in the same way as the one in Figure 5.34. Note that there is very little correlation in this image, and specifically almost no column (anti-)correlations. Even the hysteresis stripe is very suppressed.

reasonable to expect something like that. Here, we see that the effect, while small, is measurable over almost the entire width of the image.[19]

For comparison, we also show Figure 5.35, the correlation image for a dark frame from a Hamamatsu ImagEM EMCCD camera. We see that there is almost no correlation across the image here, even compared to the ProEM. We also note that the hysteresis streak is much smaller in size, but still present.

[19] Note that the hysteresis correlation extends both left and right from the origin. This means that each pixel is slightly correlated with both the previous pixels read *and* the pixels that have yet to be read. It is tempting to think that this implies the readout amplifier can somehow be affected by the values that it will read some small amount of time in the future. Sadly, this is not a way to see into the future which we could use to make a bundle of money fighting crime before it happens. It is simply a result of the symmetry inherent in correlation functions. If the second pixel the amplifier reads is skewed slightly towards the value of the first, then the two are correlated. The first pixel value affected the second (and not the reverse), but the correlation function, just being the product of the two values, simply says that the two are correlated, regardless of whether one pixel affected the other, or vice versa.

CHAPTER VI

CONCLUSION

We have successfully demonstrated an all-optical one-way barrier [18, 20, 21]. There were some unexpected obstacles that we encountered, which we worked around, and for which we developed ideas to avoid for future experimenters. We also demonstrated that the barrier can be modeled with a rather simplistic computer simulation, and explicitly calculated that barriers of this sort do not violate the second law of thermodynamics.

We also developed what we believe is a novel model of the noise present on electron-multiplying charge-coupled device cameras (EMCCD cameras). While the underlying assumptions were standard, we carried the assumptions through to a model that accurately predicts the histograms of these cameras. In one limit, our model is essentially identical to some photomultiplier models, and produces the same exponential formula for the output statistics [103, 104]. Our model provides what we believe to be a more accurate method to test certain noise measurements of these types of cameras than any of the methods currently in use today. We then used our model to investigate and compare several of the EMCCD cameras on the market today [101, 102]. As a result of this comparison, we were able to select what we believe to be the best of these cameras for our future experiments probing theories of quantum measurements.

APPENDIX A

A BASIC THEORY OF ELECTRON-MULTIPLYING CHARGE-COUPLED DEVICES

This chapter starts with the basic model of an electron-multiplying charge-coupled device (EMCCD), as described in Section V.1 and Section V.3. We then use that model to develop an in-depth description of the noise statistics of an EMCCD. In our model, electrons are multiplied in a random fashion as they are shifted through the multiplying stage, similar to how they are multiplied within a photomultiplier. As a result, in one limit, our formula reproduces one form of the output distribution for photomultipliers [103, 104].

A Brief Overview of EMCCD Operation

In this appendix we will compute actual noise statistics and model histograms for EMCCDs. To do so, we will adopt the following model on how an EMCCD works, outlined by Figure A.1, based on generally-available descriptions of how EMCCDs work [38, 40]. The model is that the signal, thermal noise, and CIC produce some cell-charge distribution that enters the EM stage of an EMCCD. The EMCCD amplifies the distribution by stretching is out, resulting in a net gain but with much overlap, and finally the readout electronics add some Gaussian noise to the signal, before the signal is read out by noise-less electronics.

We first deal with a single cell on a CCD. Before the exposure begins, the charge in a cell is the sum of a small amount of thermal noise, and CIC from the row being shifted around. During the exposure, the cell gains a certain amount of charge from more thermal noise, and, independently, a signal from the amount of light shining on the cell. After the exposure, more CIC is added due to more row-shifting, and a tiny amount of thermal noise is added. The CIC and thermal signals are both independent, Poissonian variables, and so, as mentioned in Section B.4, their sum

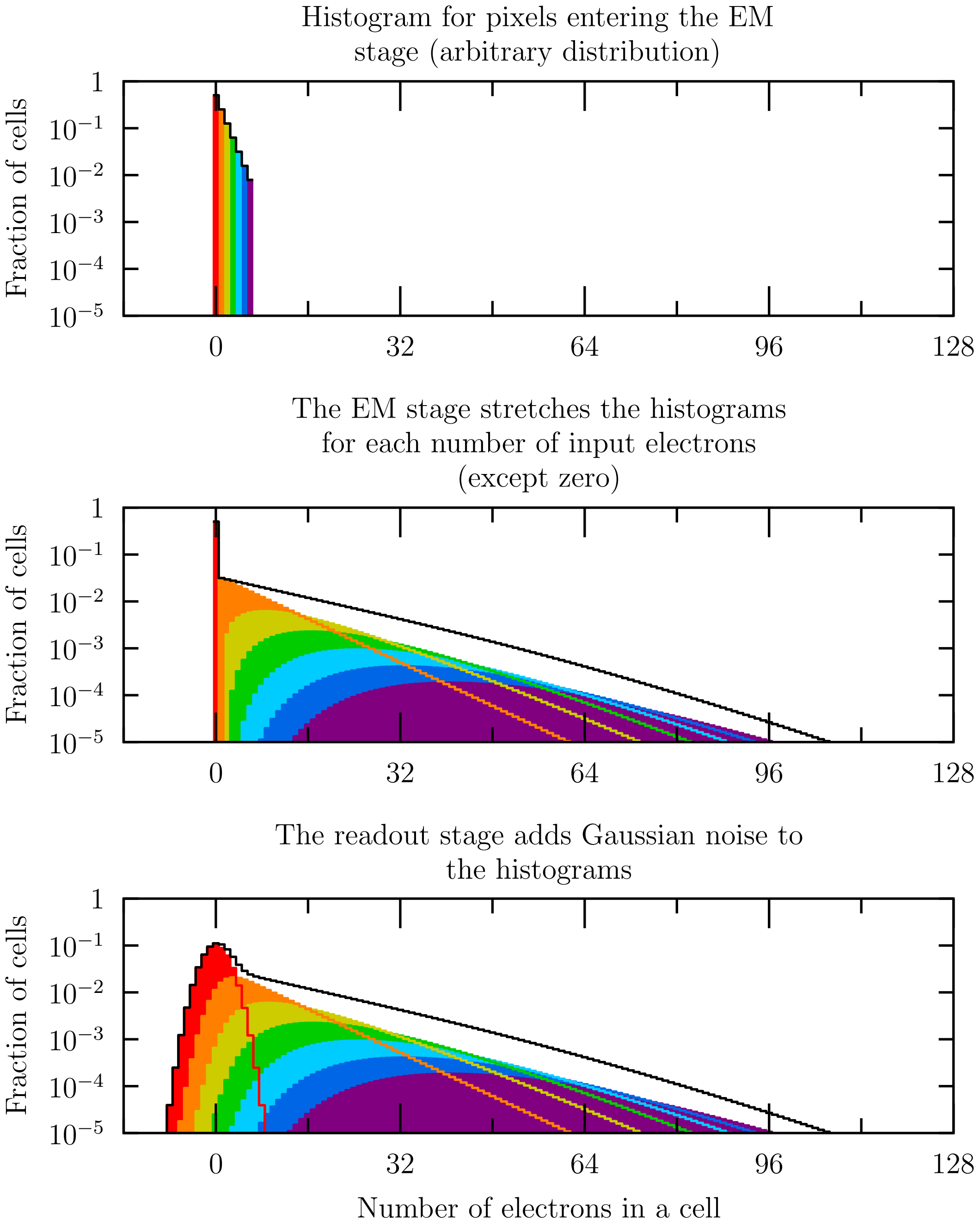


Figure A.1. An example of how the EMCCD histogram changes through the steps of our EMCCD model. The black curve shows the actual total histogram, while the colors demonstrate how individual numbers of electrons are transformed. This parameters for this figure were chosen to illustrate the progression; see Figure 5.5 for a similar figure with parameters chosen to match the cameras we investigated.

is also a Poissonian variable, with a mean and variance that's the sum of the means and variances of the individual variables. Thus, the charge in the cell as it enters the EM stage is given by:

$$\text{charge} = \text{signal} + \text{Poissonian noise}$$

The cell-charge distribution for the signal is determined by the signal. For a faint light source, it is also Poissonian, but we will not be using that yet. The cell-charge distribution for the noise is a Poissonian with a mean and variance of σ_{CIC}. Technically speaking, σ_{CIC} is the sum of the variances of the thermal noise (which is proportional to the total time of the exposure plus the time it took to move the charge around) and the CIC noise, which is proportional to the number of rows on the CCD. However, only the combined quantity σ_{CIC} matters, and since the CIC will be the dominant component for the short exposures we plan to use, we will just use the symbol σ_{CIC}, and include any thermal contributions in it. Although the amount of CIC on a given cell should be position independent, there may be some position dependence. Ideally, every cell on a CCD undergoes the same number of vertical shifts before readout. The n^{th} row of an N-row CCD accumulates n vertical shifts of CIC before the exposure begins, as the CCD is cleared of charge by performing vertical shifts. After the exposure, the n^{th} row of the CCD accumulates another $N - n$ vertical shifts of CIC. Since each vertical shift basically adds another independent Poissonian distribution, the result is the same as adding N vertical shifts of CIC to every cell on the CCD. However, there may be deviations to this. For example, if the clock signal is not perfectly evenly distributed across the CCD, or the horizontal shifts do contribute to CIC, there will be position dependence to σ_{CIC}. This can be modeled by assuming that σ_{CIC} is dependent on the cell position, but as that effect tends to be small, and we will ignore it.

The charge from the cell is eventually shifted to the EM stage, where the number of electrons is randomly multiplied. As described in Section V.3, with each step, there is a small probability that any given electron will be doubled as it travels to the next step. This is repeated many times (however many cells are in the EM stage).

After the EM stage, the charges pass through the readout amplifier, which effectively adds a small amount of Gaussian noise to the cell before being converted to a number in memory. We will attempt to describe this process mathematically.

Effect of the EM Stage on the Mean and Variance of the Number of Electrons in a Cell

Let N_i be the number of electrons in the i^{th} cell of the EM stage, and let $\mathcal{P}_i(N)$ be the probability distribution for that number. Assume that as the charges are moved from the i^{th} cell to the $(i+1)^{\text{th}}$ cell, each charge has a small probability g to excite a valence electron to the conduction band. g is small, so we can safely assume the electrons will not generate a second electron (indeed, they may not have the energy to excite a second electron). We further assume that each possible excitation is independent of others. Under these assumptions, the excited electrons will have a binomial distribution, where each of the N_i electrons has a probability of g of producing a new electron. This allows us to write:

$$N_{i+1} = N_i + \Delta N_i \tag{A.1}$$

Given N_i, ΔN_i has a binomial distribution for N_i events, each with a probability of g, so, by Equation (B.30) and Equation (B.31):

$$\begin{aligned} \text{mean}(\Delta N_i) &= gN_i \\ \text{var}(\Delta N_i) &= kN_i \\ k &:= g\,(1-g) \end{aligned}$$

In the limit of very small g, the binomial distribution becomes well-approximated by a Poissonian distribution, as mentioned in Section B.5. Furthermore, if we assume electrons can excite multiple electrons, we might assume that ΔN_i has a Poissonian distribution with mean gN_i. In either case, the above equations hold if we just replace k with g instead of $g\,(1-g)$.

We have stated that there is negligible CIC contribution in a horizontal clock. However, this is a gain stage where higher voltages are present to help excite extra electrons, and so it is possible that extra electrons will be generated, possibly from thermal excitation enhanced by the voltages. This would be more analogous to the dark current in a photomultiplier than the normal CIC. Thus, we will add a Poissonian random variable with mean l (and therefore variance l) to ΔN_i, to represent this extra leakage (Section V.5 discusses why we include this term and what it represents in greater detail). We choose to add this quantity *after* the gain. If we choose to add this quantity before the gain, it will affect the amount of amplification because it increases the mean number of electrons. In reality, the two effects probably intertwine, but in the limit of many steps, each with small amounts of gain and leakage, the alternation of gain and leakage should model reality well. In fact, as we will see in Equation (A.25), the order of these operations does not matter as the number of steps in the EM stage becomes large. This addition does not help in solving the equations for the mean and variance of the distribution; indeed, the equations for the $l = 0$ case are much easier to derive if we start without the l terms, but we will do the derivation with them. Since the l term is independent, it just adds to the mean and variance, so we can write the mean and variance of ΔN_i for a given N_i:

$$\mathrm{mean}(\Delta N_i) = gN_i + l \tag{A.2}$$

$$\mathrm{var}(\Delta N_i) = kN_i + l \tag{A.3}$$

$$k := g\,(1 - g)$$

Once again, if we would rather assume a Poissonian distribution for the electron-multiplication, we can replace k with g. In fact, in a continuous limit, where $g \to 0$ and the number of steps goes to ∞, the binomial distribution becomes Poissonian, and k equals the Poissonian value of g to first order in g.

Our next step is to write the mean and variance of $N_{i+1} = N_i + \Delta N_i$. The means just add (Equation (B.4)), but the variances do not just add, because ΔN_i is *not* independent of N_i (the mean of ΔN_i, for instance, is dependent on N_i), so we need to use the full version of Equation (B.5). This requires the variances of N_i and ΔN_i, as well as the covariance of N_i and ΔN_i. To compute these quantities, let $\Delta N_i = b + p$, where b is the multiplied electron gain term (with either a binomial distribution or a Poisson distribution), and p is the leakage term (with a Poisson distribution). We already know that $\text{mean}(b) = gN_i$, $\text{var}(b) = kN_i$, and $\text{mean}(p) = \text{var}(p) = l$, which agrees with Equation (A.2) and Equation (A.3). The problem is that the expectation values in those equations assume a fixed N_i. If N_i is itself a variable with some distribution, then that changes the expectation values. Here is how to compute an expectation value if N_i is not a fixed value:

$$\langle N_i \Delta N_i \rangle = \langle bN_i \rangle + \langle pN_i \rangle$$

p is independent of N_i, so we can write $\langle pN_i \rangle = \langle p \rangle \langle N_i \rangle$, and use $\langle p \rangle = l$. However, b and N_i are not independent, so we need to perform the weighted sum of bN_i:

$$= \sum_{b,N_i} \mathcal{P}\Big(b \bigcap N_i\Big) bN_i + l \langle N_i \rangle$$

We know the mean and variance of b for a given N_i, so we choose to rewrite $\mathcal{P}(b \bigcap N_i)$ in terms of a conditional probability, using Equation (B.10):

$$\begin{aligned} &= \sum_{b,N_i} \mathcal{P}(b \mid N_i) \mathcal{P}_i(N_i) bN_i + l \langle N_i \rangle \\ &= \sum_{N_i} \mathcal{P}_i(N_i) N_i \left(\sum_{b} \mathcal{P}(b \mid N_i) b \right) + l \langle N_i \rangle \end{aligned}$$

The sum in parentheses is the mean of b for a given N_i, which we know is gN_i. With that substitution, the remaining sum will become an average of N_i^2.

$$\begin{aligned} &= \sum_{N_i} \mathcal{P}_i(N_i) N_i g N_i + l \langle N_i \rangle \\ &= g \sum_{N_i} \mathcal{P}_i(N_i) N_i^2 + l \langle N_i \rangle \\ &= g \left\langle N_i^2 \right\rangle + l \langle N_i \rangle \end{aligned}$$

The trick used to evaluate $\langle bN_i \rangle$ was basically to evaluate the expectation over all b for a given N_i, and then evaluate over all N_i. We can use this trick for the other expectation values we need (remember the p is independent of b and N_i):

$$\begin{aligned}
\langle \Delta N_i \rangle &= \langle b \rangle + \langle p \rangle \\
&= \sum_{b,N_i} \mathcal{P}\Big(b \bigcap N_i\Big) b + l \\
&= \sum_{N_i} \mathcal{P}_i(N_i) \left(\sum_b \mathcal{P}(b \mid N_i) b \right) + l \\
&= \sum_{N_i} \mathcal{P}_i(N_i) g N_i + l \\
&= g \langle N_i \rangle + l \\
\left\langle (\Delta N_i)^2 \right\rangle &= \left\langle b^2 + 2bp + p^2 \right\rangle \\
&= \left\langle b^2 \right\rangle + 2 \langle b \rangle \langle p \rangle + \left\langle p^2 \right\rangle \\
&= \sum_{b,N_i} \mathcal{P}\Big(b \bigcap N_i\Big) b^2 + 2 \langle p \rangle \sum_{b,N_i} \mathcal{P}\Big(b \bigcap N_i\Big) b + \text{var}(p) + \langle p \rangle^2 \\
&= \sum_{N_i} \mathcal{P}_i(N_i) \left(\sum_b \mathcal{P}(b \mid N_i) b^2 \right) \\
&\quad +2l \sum_{N_i} \mathcal{P}_i(N_i) \left(\sum_b \mathcal{P}(b \mid N_i) b \right) + l + l^2 \\
&= \sum_{N_i} \mathcal{P}_i(N_i) \left(\text{var}(b) + \text{mean}(b)^2 \right) + 2 \langle p \rangle \sum_{N_i} \mathcal{P}_i(N_i) \left(\text{mean}(b) \right) + l + l^2 \\
&= \sum_{N_i} \mathcal{P}_i(N_i) \left(k N_i + g^2 N_i^2 \right) + 2l \sum_{N_i} \mathcal{P}_i(N_i) \left(g N_i \right) + l + l^2 \\
&= \sum_{N_i} \mathcal{P}_i(N_i) \left(k N_i + g^2 N_i^2 \right) + 2l \sum_{N_i} \mathcal{P}_i(N_i) \left(g N_i \right) + l + l^2 \\
&= k \langle N_i \rangle + g^2 \left\langle N_i^2 \right\rangle + 2lg \langle N_i \rangle + l + l^2
\end{aligned}$$

With these values, we can write out the mean and variance of N_{i+1}:

$$\begin{aligned}
\text{mean}(N_{i+1}) &= \langle N_i \rangle + \langle \Delta N_i \rangle \\
&= (\text{mean}(N_i)) + g\,(\text{mean}(N_i)) + l \\
&= (1+g)\,(\text{mean}(N_i)) + l && \text{(A.4)} \\
\text{var}(N_{i+1}) &= \left\langle (N_i + \Delta N_i)^2 \right\rangle - \langle (N_i + \Delta N_i) \rangle^2 \\
&= \left\langle N_i^2 \right\rangle + 2\,\langle N_i \Delta N_i \rangle + \left\langle (\Delta N_i)^2 \right\rangle - \langle N_i \rangle^2 - 2\,\langle N_i \rangle\,\langle \Delta N_i \rangle - \langle \Delta N_i \rangle^2 \\
&= \left\langle N_i^2 \right\rangle + 2\left(g \left\langle N_i^2 \right\rangle + l\,\langle N_i \rangle \right) + \\
&\quad k\,\langle N_i \rangle + g^2 \left\langle N_i^2 \right\rangle + 2lg\,\langle N_i \rangle + l + l^2 - \\
&\quad \langle N_i \rangle^2 - 2\,\langle N_i \rangle\,(g\,\langle N_i \rangle + l) - (g\,\langle N_i \rangle + l)^2 \\
&= (1+g)^2 \left(\left\langle N_i^2 \right\rangle - \langle N_i \rangle^2 \right) + k\,\langle N_i \rangle + l \\
&= (1+g)^2\,\text{var}(N_i) + k\text{mean}(N_i) + l && \text{(A.5)}
\end{aligned}$$

Now we need only solve the recurrence relations in Equation (A.4) and Equation (A.5). Equation (A.4) is simpler. Starting with mean(N_0), and using Equation (A.4) to write out mean(N_1), mean(N_2), and so on, we can very quickly see:

$$\begin{aligned}
\text{mean}(N_i) &= (1+g)^i\,(\text{mean}(N_0)) + l \sum_{j=0}^{i-1} (1+g)^j \\
&= (1+g)^i\,(\text{mean}(N_0)) + l\frac{(1+g)^i - 1}{g} && \text{(A.6)}
\end{aligned}$$

where the sum is assumed to be 0 when $i = 0$, and we explicitly solved it using a formula for a geometric series. It is fairly simple to verify that either form of Equation (A.6) satisfies both the base case mean(N_0) = mean(N_0) and the recursion relation in Equation (A.4). The variance is trickier, but can be solved in a similar manner. It is a little easier to use this form of Equation (A.6):

$$\text{mean}(N_i) = (1+g)^i \left(\text{mean}(N_0) + \frac{l}{g} \right) - \frac{l}{g}$$

Start with var(N_0), and use Equation (A.5) to write out var(N_1) in terms of mean(N_0) and var(N_0) (use the above form of Equation (A.6) for mean(N_0)). We can then use Equation (A.5) (and the above form of Equation (A.6)) to write out var(N_2) in terms of mean(N_0) and var(N_0). Continuing a few more terms, if we always collect the terms into a var(N_0) term, a $k\,(\text{mean}(N_0) + l/g)$ term, and a $l\,(1 - k/g)$ term, we see that the var(N_0) term is a geometric sequence, and all the others are geometric series (or a geometric series times a geometric sequence). Again, it is quite easy to show that this form solves both the var(N_0) = var(N_0) base condition and the recurrence relation in Equation (A.5):

$$\begin{aligned}
\text{var}(N_i) = \;& (1+g)^{2i}\,\text{var}(N_0) + \\
& (1+g)^{i-1}\,k\left(\text{mean}(N_0) + \frac{l}{g}\right)\sum_{j=0}^{i-1}(1+g)^j + \\
& l\left(1 - \frac{k}{g}\right)\sum_{j=0}^{i-1}(1+g)^{2j} \\
= \;& (1+g)^{2i}\,\text{var}(N_0) + \\
& (1+g)^{i-1}\,k\left(\text{mean}(N_0) + \frac{l}{g}\right)\frac{(1+g)^i - 1}{g} \\
& l\left(1 - \frac{k}{g}\right)\frac{(1+g)^{2i} - 1}{g\,(2+g)}
\end{aligned} \tag{A.7}$$

where we substituted an analytic form for each of the geometric series again (and simplified), and we state that all the sums (which includes the entire factor inside the braces) are 0 at $i = 0$.

We can now bring everything together. Assume the actual signal has a mean M and variance V, and the gain stage has n steps each with gain g. The charge entering the EM stage is the sum of two presumably independent values, the signal and the CIC noise (including thermal noise). Since the are independent, the means and variances simply add:

$$\begin{aligned}
\text{mean}(N_0) &= M + \sigma_{\text{CIC}} \\
\text{var}(N_0) &= V + \sigma_{\text{CIC}}
\end{aligned}$$

Equation (A.6) and Equation (A.7) can propagate these through the EM stage:

$$\begin{aligned}
\text{mean}(N_n) &= (1+g)^n\left(M+\sigma_{\text{CIC}}\right)+l\frac{(1+g)^n-1}{g}\\
\text{var}(N_n) &= (1+g)^{2n}\left(V+\sigma_{\text{CIC}}\right)+\\
&\quad (1+g)^{n-1}k\left(M+\sigma_{\text{CIC}}+\frac{l}{g}\right)+\frac{(1+g)^n-1}{g}\\
&\quad l\left(1-\frac{k}{g}\right)\frac{(1+g)^{2n}-1}{g\left(2+g\right)}
\end{aligned}$$

We realize that the signal is multiplied by the factor $(1+g)^n$, which we define as the gain G of the EM stage. We also choose to replace l with L, where we define L to be the expected mean value coming out of the gain stage with a gain of unity, and no input. We find that by taking the limit of the l term in the above mean as $g \to 0$, which happens to be the same limit as the derivative of lx^n at $x=1$, which is nl. Note that this is exactly what we should expect, since if there is no gain, and we go through n cells which each add an independent amount of noise with mean l, we would expect a total of nl, on average. Thus, we define the total leakage as:

$$L := nl \tag{A.8}$$

Finally, the readout stage adds some Gaussian noise, with a mean m_r and standard deviation σ_{r}. Since it is independent of the signal, the means and variances add. The result is the mean and variance of the distribution of the output of the EMCCD for a given signal mean and variance, assuming a gain of G:

$$\text{mean(output signal)} = GM+G\sigma_{\text{CIC}}+(G-1)\frac{L}{ng}+m_r \tag{A.9}$$

$$\begin{aligned}
\text{var(output signal)} &= G^2\left(V+\sigma_{\text{CIC}}\right)+ \qquad \text{(A.10)}\\
&\quad \frac{G\left(G-1\right)}{(1+g)}\frac{k}{g}\left(M+\sigma_{\text{CIC}}+\frac{L}{ng}\right)+\\
&\quad \frac{G^2-1}{(2+g)}\frac{L}{ng}\left(1-\frac{k}{g}\right)+\sigma_{\text{r}}\\
k &:= \begin{cases} g & \text{for Poissonian EM gain}\\ g\left(1-g\right) & \text{for binomial EM gain}\end{cases}\\
g &:= G^{\frac{1}{n}}-1
\end{aligned}$$

Here, we have defined g in terms of the total gain G by solving $G = (1+g)^n$ for g.

This equation can be simplified with some reasonable assumptions. We have mentioned that EM gains (G) are typically around 1000, and the number of steps (n) is typically around 500. With these two numbers, g is between 1% and 2%, so it makes sense to try a small g approximation [38, 101, 102]. Taylor expanding the definition of g used for Equation (A.9) in $1/n$ gives $g \approx \ln(G)/n$. We can also arrive at this by taking the limit of ng as $n \to \infty$. We used the trick of writing $ng = \left(G^{1/n} - 1\right) / (1/n)$, and then using l'Hôpital's rule (presented in typical differential calculus courses) for evaluating a limit of the form 0/0. For a maximum EM gain of $G \lesssim 1000$, and at least $n \gtrsim 500$ steps, we can approximate with $g \approx 0$ and $ng \approx \ln(G)$, with a maximum error on the order of 1%.

$$g = \frac{\ln(G)}{n} \text{ as } n \to \infty \tag{A.11}$$

Since the linearity of the readout amplifiers, according to several camera data sheets, is at best around 1%, and the EM gain stage probably varies by about this much from step to step, this amount of error is probably unavoidable [102]. Note that, with such limits, $k/g \to 1$, regardless of which version of k we used. This is because a binomial distribution approaches a Poissonian distribution in the large n limit, as mentioned in Section B.5. We now have a continuous approximation for the mean and variance of the signal:

$$\text{mean(output signal)} = GM + G\sigma_{\text{CIC}} + (G-1)\,L_G + m_r \tag{A.12}$$

$$\begin{aligned} \text{var(output signal)} = \; & G^2\,(V + \sigma_{\text{CIC}}) + \\ & G\,(G-1)\,(M + \sigma_{\text{CIC}} + L_G) + \sigma_{\text{r}} \end{aligned} \tag{A.13}$$

$$L_G := \frac{L}{\ln(G)}$$

We are now in a position to explain Equation (V.3) and Equation (V.4) more fully. First, CCD manufacturers never quote a value for L or L_G. In fact, in some private communications with some employees of a company that makes EMCCD cameras, we got the impression that EMCCD camera manufacturers are unaware of this leakage effect, repeating that there is no significant clock-induced charge in any horizontal shift, including in the electron-multiplying stage. Presumably, the value of L is strongly dependent on the gain (because l probably depends strongly on the exact voltage used for each step in the EM gain stage), but we hope it is rather small, so, for now, we will set $L_G = 0$. The mean value given in Equation (A.12) is presumably GM plus known constant values. Therefore, we can reliably subtract off everything but the signal GM. The noise used in a signal-to-noise ratio is typically the standard deviation, so we just take the square root of Equation (A.13). The result, with these limits, is:

$$\begin{aligned}\mathrm{SNR} &= \frac{GM}{\sqrt{G^2\left(V+\sigma_{\mathrm{CIC}}\right)+G\left(G-1\right)\left(M+\sigma_{\mathrm{CIC}}\right)+\sigma_{\mathrm{r}}^2}} \\ &= \frac{M}{\sqrt{\left(V+\sigma_{\mathrm{CIC}}\right)+\left(1-\frac{1}{G}\right)\left(M+\sigma_{\mathrm{CIC}}\right)+\sigma_{\mathrm{r}}^2/G^2}}\end{aligned} \tag{A.14}$$

Constant intensity light has Poissonian statistics, which is a result of there being a constant rate of photon detection. Thus, we can equate $M = V$, giving the signal-to-noise ratio of:

$$\begin{aligned}\mathrm{SNR} &= \frac{GM}{\sqrt{G^2\left(M+\sigma_{\mathrm{CIC}}\right)+G\left(G-1\right)\left(M+\sigma_{\mathrm{CIC}}\right)+\sigma_{\mathrm{r}}^2}} \\ &= \frac{GM}{\sqrt{\left(2G^2-G\right)\left(M+\sigma_{\mathrm{CIC}}\right)+\sigma_{\mathrm{r}}}} \\ &= \frac{M}{\sqrt{\left(2-\frac{1}{G}\right)\left(M+\sigma_{\mathrm{CIC}}\right)+\sigma_{\mathrm{r}}^2/G^2}}\end{aligned} \tag{A.15}$$

In this derivation, we included the dark noise in σ_{CIC}. Once we note that, we see that this version reproduces Equation (V.4), with $f = 2 - 1/G$ (so that $f \to \sqrt{2} \approx 1.4$ for large G), and becomes Equation (V.3) in the $G \to 1$ limit (no EM gain). This explains how the EM stage adds extra noise. We also note how the EM gain also effectively suppresses the standard readout noise by a factor of the EM gain, since the only occurrence of σ_{r} in the signal-to-noise ratio is σ_{r}/G.

There is one final point we should make about Equation (A.12) and Equation (A.13). We should expect the $G \to 1$ limit to be well-defined, as then these equations should represent the addition of some extra clock-induced charge. In particular, L was defined as the amount of extra clock-induced charge that should be produced in the $G \to 1$ limit. However, if we take $G \to 1$, then $\ln(G) \to 0$, and $L_G = L/\ln(G)$ diverges, so the equations are singular in this limit. This singularity is removable, though, because everywhere L_G appears in those equations, it is multiplied by $G-1$, and the limit as $G \to 1$ of $(G-1)\,L_G = (G-1)\,L/\ln(G)$ is L, as is easily verified using l'Hôpital's rule. In particular, if we let $G = 1+\epsilon$, and let $\epsilon \to 0$, then

$$L_G = \frac{L}{\epsilon - \frac{\epsilon^2}{2} + \mathcal{O}(\epsilon^3)} = \frac{L}{\epsilon}\left(1 + \frac{\epsilon}{2} + \mathcal{O}\big(\epsilon^2\big)\right)$$

and Equation (A.12) and Equation (A.13) reduce to:

$$\begin{aligned}
\text{mean(output signal)} &= GM + G\sigma_{\text{CIC}} + (1+\epsilon-1)\,\frac{L}{\epsilon}\left(1+\frac{\epsilon}{2}+\mathcal{O}\big(\epsilon^2\big)\right) + m_r \\
&= M\,(1+\epsilon) + \sigma_{\text{CIC}}\,(1+\epsilon) + L\left(1+\frac{\epsilon}{2}\right) + m_r + \mathcal{O}\big(\epsilon^2\big) \\
\text{var(output signal)} &= \ G^2\,(V+\sigma_{\text{CIC}}) + \\
&\quad\ G\,(1+\epsilon-1)\left(M + \sigma_{\text{CIC}} + \frac{L}{\epsilon}\left(1+\frac{\epsilon}{2}+\mathcal{O}\big(\epsilon^2\big)\right)\right) + \sigma_{\text{r}} \\
&= \ (V+\sigma_{\text{CIC}})\,(1+2\epsilon) + \\
&\quad\ \left(\epsilon M + \epsilon\sigma_{\text{CIC}} + L\left(1+\frac{\epsilon}{2}\right)\right)(1+\epsilon) + \sigma_{\text{r}} + \mathcal{O}\big(\epsilon^2\big) \\
&= \ V\,(1+2\epsilon) + \sigma_{\text{CIC}}\,(1+3\epsilon) + \\
&\quad\ L\left(1+\frac{3\epsilon}{2}\right) + \sigma_{\text{r}} + \mathcal{O}\big(\epsilon^2\big)
\end{aligned}$$

In this form, we can see that in the $\epsilon \to 0$ limit ($G \to 1$), L does indeed just become an addition to the CIC noise, but the first-order ϵ dependence is different than that of the CIC noise. This is because the leakage gets amplified differently than the CIC noise. All of the CIC noise experiences the same gain, whereas only the leakage that occurs near the beginning of the electron-multiplying stage gets the full gain. Leakage that occurs near the end of the EM stage experiences no gain. In the low-gain limit, we see that the leakage experiences the *average* of those two gain extremes, getting approximately half the increase in effect as that of the CIC noise. At higher gains, where there is a significant probability of an electron being multiplied more than once, the gain becomes nonlinear, and while the leakage and CIC experience different gains, the ratio is no longer one-half.

That leaves us only with the question as to why our formula has a singularity in this limit.

Generating Histograms for EMCCDs

Using this EMCCD model, we can actually generate histograms for EMCCD data, including all the noise sources. In some limits, we can even compute analytic results. We will use the same model described in Section V.3, using the notation introduced there and in Subsection A.2.

As already described, the cell-charge histogram before the EM stage (N_0) results from the sum of the signal and the combined CIC and thermal noise, which we abbreviate as σ_{CIC}. As described in Section B.1, since the signal and noise are independent, the cell-charge histogram is the convolution of the signal and noise histograms:

$$\mathcal{H}\{N_0\} = \mathcal{H}\{\text{signal}\} * \mathcal{H}\{\text{CIC}\} \tag{A.16}$$

$\mathcal{H}\{\text{CIC}\}$ is a Poissonian distribution with mean σ_{CIC}.

The tricky part is generating the histogram for N_{i+1} from the histogram for N_i, since the distribution of added electrons in each step of the histogram is dependent on the number of electrons already present. There is a tempting trick one could try to use to infer what the histogram should look like: Look at the logarithm of the number of electrons in each cell. We start with some number of electrons in the first cell, N_0. Subsequent cells have, on average, $(1+g)$ times as many electrons as the initial cell, so we could write the number of electrons in cell n as

$$N_n = N_0 \prod_{i=0}^{n-1} (1+g_i)$$

where g_i is some random variable with mean g. We will show shortly that that is actually the correct mean. Now, if we take the logarithm of this, we find:

$$\ln(N_n) = \ln(N_0) + \sum_{i=0}^{n-1} \ln(1+g_i)$$

If we treat $\ln(1+g_i)$ as independent random variables with identical distributions, then we can immediately apply the central limit theorem and say that, in the limit of large n, the distribution of $\ln(N_n)$ is Gaussian, and we can infer the mean and standard deviation of that from the mean and standard deviation of each step. Thus, if the above argument were correct, the distribution of N_n would be *log-normal*, which is defined as the distribution of a variable whose logarithm has a normal (Gaussian) distribution. These distributions occur when some random, independent multiplying factor is repeatedly applied to the initial quantity, in which case, the above argument holds.

The above argument fails because the random multiplying factor is *not* independent of the previous factors. More specifically, it is not independent of the current number of electrons, which could be dependent on the initial number, or the previous multiplying factors applied. This happens because of our assumption that the probability of adding extra electrons is independent of the presence of other electrons. For each electron that passes through a cell, there is a certain distribution for adding electrons, with some variance. Since each addition is independent, the sum over N_i electrons gives a distribution with a variance that is N_i times that original variance. This is reflected in Equation (A.2) and Equation (A.3) with $l = 0$, since we are ignoring leakage for now. In those equations, k is the variance of ΔN_i for a single electrons (the case where $N_i = 1$), and the variance of adding up N_i such variables is N_i times k. Now, let us look at the multiplying factor $(1 + g_i)$ from our above argument:

$$\begin{aligned}
(1+g_i) &:= \frac{N_{i+1}}{N_i} = \frac{N_i + \Delta N_i}{N_i} = 1 + \frac{\Delta N_i}{N_i} \\
\text{mean}(1+g_i) &= \text{mean}\left(1 + \frac{\Delta N_i}{N_i}\right) = 1 + \frac{gN_i}{N_i} = 1 + g \\
\text{var}(1+g_i) &= \text{var}\left(1 + \frac{\Delta N_i}{N_i}\right) = \text{var}\left(\frac{\Delta N_i}{N_i}\right) = \frac{\text{var}(\Delta)N_i}{N_i^2} = \frac{kN_i}{N_i^2} = \frac{k}{N_i}
\end{aligned}$$

Here, we have computed the means and variances assuming that N_i is fixed, and only ΔN_i is a random variable, allowing us to use Equation (A.2) and Equation (A.3) with $l = 0$. Here, we see that the mean of g_i is indeed g, as claimed above, but the variance of it is *not* constant. Thus, the distributions of the g_i for each stage is dependent on what happened before, so the central limit theorem applied to the sum of the logarithms is not fully applicable. However, as the number of electrons becomes rather large, the variance computed above flattens out as $1/N_i$, and so the distributions do not change rapidly and the central limit theorem provides a very rough approximation. Thus, the final histogram will have some resemblance to a log-normal distribution.

We now begin a careful derivation of the final histogram given our assumptions on how an EMCCD works, and will shortly make a further assumption that the EM stage is continuous, where the number of stages is infinite, with each stage individually adding almost no gain. We know the distribution for the number of electrons in cell $i+1$ (N_{i+1}) for a given number of electrons in cell i (N_i) is binomial (or Poissonian) with mean $(1+g)\,N_i$, plus an independent Poissonian leakage which we will add shortly. The chances of getting n electrons in the $(i+1)^{\text{th}}$ cell conditioned on there begin m electrons in the i^{th} cell is the probability of gaining $n-m$ electrons given that there are already m electrons. The total probability of getting n electrons in the $(i+1)^{\text{th}}$ cell is the conditioned probability multiplied by the probability of having m electrons in the i^{th} cell, summed over all possible m. This is basically Equation (B.12) applied to the EM stage, and looks like a convolution of the cell-charge histogram for N_i with the distribution for ΔN_i (but, as in Equation (B.13), it is not exactly a convolution as one of the functions depends on the m as well as $n-m$):

$$\mathcal{H}\{N_{i+1}\}(n) = \sum_{m=0}^{n} \mathcal{H}\{N_i\}(m)\mathcal{G}(n-m \mid m) \text{ without leakage}$$

Here, we have used $\mathcal{G}(n-m \mid m)$ as the probability of gaining $n-m$ electrons from one cell in the EM stage to the next given that the first cell had m electrons. The m sum ends at n because we assume the EM stage does not lose electrons, so the only way to get n electrons in the $(i+1)^{\text{th}}$ stage is if the i^{th} stage had at most n electrons. Depending on your assumptions on how the EM gain works, this would be approximately a binomial distribution or a near-Poissonian distribution with mean gm. We will eventually assume the continuous limit again, where $g \to 0$, and $\mathcal{G}(n-m \mid m)$ becomes very nearly Poissonian, as discussed in Section B.4. Adding leakage is straightforward. As previously discussed, the amount of leakage is assumed to be independent of the number of electrons in the cell, and has a Poissonian distribution with mean l. In our model, we assume the leakage happens after the gain for any given step, so the actual cell-charge histogram is a convolution of the no-leakage cell-charge histogram above with the Poissonian leakage distribution $\mathcal{L}(n)$:

$$\begin{aligned}\mathcal{H}\{N_{i+1}\}(n) &= \sum_{m'=0}^{n} \mathcal{L}(n-m') \left\{ \sum_{m=0}^{m'} \mathcal{H}\{N_i\}(m) \mathcal{G}(m'-m \mid m) \right\} \\ &= \sum_{m=0}^{\infty} \left\{ \sum_{m'=0}^{\infty} \mathcal{L}(n-m') \mathcal{G}(m'-m \mid m) \right\} \mathcal{H}\{N_i\}(m) \qquad \text{(A.17)}\end{aligned}$$

Again, the sums terminate because we assume the EM stage does not lose electrons. However, in re-ordering the sums, we take the upper limits to ∞, and incorporate the upper limits in $\mathcal{G}$ and $\mathcal{L}$ by defining them to be zero if the argument is negative.

Note that Equation (A.17) is linear in $\mathcal{H}\{N_i\}$, and is in the form of a matrix multiplication. Since we are assuming that the EM stage does not lose electrons, we can define $\mathcal{G}(n \mid m)$ and $\mathcal{L}(n)$ to be 0 whenever $n < 0$, and we need not even consider the $m < 0$ case, since we never have fewer then 0 electrons in a cell. We can therefore define a gain transfer matrix $\widetilde{\mathcal{G}}$ that, when multiplied by a cell-charge histogram for cell i, gives the cell-charge histogram for cell $i+1$ before we add leakage (or in the $l \to 0$ limit). We choose to represent $\widetilde{\mathcal{G}}$ as the usual two-dimensional matrix, except we index both the rows and columns starting from 0 instead of the more usual 1. This way, we can represent a cell-charge histogram $\mathcal{H}\{N\}(n)$ as a column vector, where row n (starting from 0) is the probability of having n electrons, and the element in row n, column m of the gain matrix $\widetilde{\mathcal{G}}$ is the probability of gaining n electrons given that there are already m electrons:

$$\widetilde{\mathcal{G}}(n, m) = \begin{cases} 0 & n < m \\ \mathcal{G}(n-m \mid m) & n \geq m \end{cases} \text{ for } n \geq 0 \text{ and } m \geq 0 \qquad \text{(A.18)}$$

Again, $\mathcal{G}(n-m \mid m)$ is either a binomial or a near-Poissonian distribution with mean gm, depending on which model of the EM stage is being used. For a binomial distribution, each of the m electrons has a g chance of being doubled, so the expected number of new electrons is gm.

Likewise, we define a leakage transfer matrix $\widetilde{\mathcal{L}}$ that, when multiplied by a cell-charge distribution for cell i after being multiplied by the gain transfer matrix, gives the cell-charge distribution for cell $i+1$. The elements of this matrix are:

$$\widetilde{\mathcal{L}}(n,m) = \begin{cases} 0 & n < m \\ \mathcal{L}(n-m) & n \geq m \end{cases} \quad \text{for } n \geq 0 \text{ and } m \geq 0 \tag{A.19}$$

where $\mathcal{L}(n-m)$ is a Poissonian distribution with mean l. The probability of *gaining* a certain number of electrons is a Poissonian distribution with a mean of gm. The *gain* in electrons if the $(i+1)^{\text{th}}$ has n electrons and the i^{th} cell had m electrons is $n-m$. This gain is what has a Poissonian distribution, which is why the argument of $\mathcal{L}(n-m)$ in Equation (A.19) is $n-m$.

With $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{L}}$, we can now very easily write out the cell-charge histogram at the end of the EM stage. Given a cell-charge distribution for cell i, we first multiply by the $\widetilde{\mathcal{G}}$, and then by $\widetilde{\mathcal{L}}$ to get the cell-charge distribution for cell $i+1$. This is equivalent to Equation (A.17), which we rearranged to show that matrix multiplication is associative, which means we can simply multiply the cell-charge distribution for cell i by $\widetilde{\mathcal{L}}\widetilde{\mathcal{G}}$ to get the cell-charge distribution for cell $i+1$. Since $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{G}}$ are constant (if we assume the leakage and gain are constant across the entire EM stage), we can easily write out the last cell-charge distribution for the EM stage:

$$\begin{aligned} \mathcal{H}\{N_n\} &= \left(\widetilde{\mathcal{L}}\widetilde{\mathcal{G}}\right)^n \mathcal{H}\{N_0\} \\ &= \left(\widetilde{\mathcal{L}}\widetilde{\mathcal{G}}\right)^n (\mathcal{H}\{\text{signal}\} * \mathcal{H}\{\text{CIC}\}) \end{aligned} \tag{A.20}$$

where we used Equation (A.16) for the histogram entering the EM stage. The final output histogram is simply the convolution of this with the readout noise Gaussian.

The elements of $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{L}}$ can be computed from Equation (A.18). Some elements are given here for reference:

$$\widetilde{\mathcal{G}} = \begin{cases} \begin{bmatrix} 1 & 0 & 0 & 0 & \\ 0 & \binom{1}{0}(1-g) & 0 & 0 & \\ 0 & \binom{1}{1}g & \binom{2}{0}(1-g)^2 & 0 & \cdots \\ 0 & 0 & \binom{2}{1}g(1-g) & \binom{3}{0}(1-g)^3 & \\ & & \vdots & & \ddots \end{bmatrix} & \text{for binomial EM gain} \\ \begin{bmatrix} 1 & 0 & 0 & 0 & \\ 0 & \frac{(g)^0}{0!}e^{-g} & 0 & 0 & \\ 0 & \frac{(g)^1}{1!}e^{-g} & \frac{(2g)^0}{0!}e^{-2g} & 0 & \cdots \\ 0 & \frac{(g)^2}{2!}e^{-g} & \frac{(2g)^1}{1!}e^{-2g} & \frac{(3g)^0}{0!}e^{-3g} & \\ & & \vdots & & \ddots \end{bmatrix} & \text{for Poissonian EM gain} \end{cases} \tag{A.21}$$

$$\widetilde{\mathcal{L}} = \begin{bmatrix} le^{-l} & 0 & 0 & 0 & \\ le^{-l} & le^{-l} & 0 & 0 & \\ \frac{l^2}{2!}e^{-l} & le^{-l} & le^{-l} & 0 & \cdots \\ \frac{l^3}{3!}e^{-l} & \frac{l^2}{2!}e^{-l} & le^{-l} & le^{-l} & \\ & & \vdots & & \ddots \end{bmatrix} \tag{A.22}$$

We can actually reduce Equation (A.20) to a relatively simple analytic form in the continuous limit. As before, $g \to 0$ and $l \to 0$, and it is only the terms ng and

nl that remain nonzero. We can expand $\widetilde{\mathcal{G}}$ and $\widetilde{\mathcal{L}}$ in g and l and get:

$$\widetilde{\mathcal{G}} = \widetilde{\mathbb{I}} + g \begin{bmatrix} 0 & 0 & 0 & 0 & \\ 0 & -1 & 0 & 0 & \\ 0 & 1 & -2 & 0 & \cdots \\ 0 & 0 & 2 & -3 & \\ & & \vdots & & \ddots \end{bmatrix} + \mathcal{O}\big(g^2\big) \tag{A.23}$$

$$\widetilde{\mathcal{L}} = \widetilde{\mathbb{I}} + l \begin{bmatrix} -l & 0 & 0 & 0 & \\ l & -l & 0 & 0 & \\ 0 & l & -l & 0 & \cdots \\ 0 & 0 & l & -l & \\ & & \vdots & & \ddots \end{bmatrix} + \mathcal{O}\big(l^2\big) \tag{A.24}$$

where we have used $\widetilde{\mathbb{I}}$ for the identity matrix. Here, we see that, to first order in g, it does not matter whether $\widetilde{\mathcal{G}}$ uses a binomial distribution or a Poissonian distribution. The two distributions become identical in this limit. Substituting these values into the $\left(\widetilde{\mathcal{L}}\widetilde{\mathcal{G}}\right)^n$ from Equation (A.20) yields:

$$\left(\widetilde{\mathcal{L}}\widetilde{\mathcal{G}}\right)^n = \left(\widetilde{\mathbb{I}} + g\widetilde{g} + l\widetilde{l} + \mathcal{O}\big(g^2, l^2, gl\big)\right)^n$$

Here, we are using $\widetilde{g}$ and $\widetilde{l}$ to represent the first-order-term matrices in Equation (A.23) and Equation (A.24), respectively. If you arithmetically expand the right hand side, you will find that the higher-order $\mathcal{O}(g^2, l^2, gl)$ terms always have higher powers of g or l than n, regardless of whether the higher-order terms commute (which they do not, in this case). Therefore, in the continuous limit, where $g \to 0$ and $l \to 0$ and only the terms composed of ng and nl remain, we can drop the higher-order terms. We can them substitute that into Equation (A.20) to get:

$$\begin{aligned} \mathcal{H}\{N_n\} &= \left(\widetilde{\mathbb{I}} + g\widetilde{g} + l\widetilde{l}\right)^n (\mathcal{H}\{\text{signal}\} * \mathcal{H}\{\text{CIC}\}) \\ n &\to \infty \end{aligned} \tag{A.25}$$

in the continuous limit. This result applies even if the order of $\widetilde{\mathcal{L}}$ and $\widetilde{\mathcal{G}}$ in the original product were reversed, supporting our claim that, in the continuous limit, the leakage and the gain parts of each step combine into a single step.[1]

We can actually compute an analytic result for the continuous limit given in Equation (A.25). To do so, we will first compute the eigenvectors of the matrix power, and then demonstrate how to decompose arbitrary initial histograms into these eigenvectors, and then sum them together after the matrix power. Observe that the eigenvectors of $g\widetilde{g} + l\widetilde{l}$ are also eigenvectors of $\widetilde{\mathbb{I}} + g\widetilde{g} + l\widetilde{l}$, and so therefore eigenvectors of the matrix power. Therefore, we will start by finding the eigenvectors of this matrix:

$$g\widetilde{g} + l\widetilde{l} = \begin{bmatrix} -l & 0 & 0 & 0 & \\ l & -g-l & 0 & 0 & \\ 0 & g+l & -2g-l & 0 & \cdots \\ 0 & 0 & 2g+l & -3g-l & \\ & & \vdots & & \ddots \end{bmatrix} =: \widetilde{\mathfrak{g}}$$

This is the sum of the two first-order matrices in Equation (A.25). $\widetilde{\mathfrak{g}}$ is lower-diagonal, with at most two non-zero elements in each row, and an enticing pattern. The eigenvalues of a lower (or upper) diagonal matrix are the diagonal elements. This is because the characteristic equation for such a matrix is just the product of the diagonal elements (once $\lambda\widetilde{\mathbb{I}}$ is subtracted). We can therefore write the eigenvalues

[1] This result also implies that, in the continuous limit, variations in the gain and leakage rate do not matter. The basic argument is to consider the quantity $Q = \prod_{i=1} n\,(1 + r_n/n)$ as $n \to \infty$, where r_n is some random number with mean 0 (and with a magnitude small enough that it cannot change the sign of the quantity). The logarithm of Q is $\ln(Q) = \sum_{i=1} n\ln(1 + r_n/n)$. As long as n is large enough, we can approximate the logarithm as a power series $\ln(1+\epsilon) = \epsilon - \epsilon^2/2 + \mathcal{O}(\epsilon^3)$. With that expansion, we find $\ln(Q) = \frac{1}{n}\sum_{i=1} nr_n + \frac{1}{n^2}\sum_{i=1} nr_n^2 + \mathcal{O}(1/n^3)$. In the large n limit, since the magnitude of r_n is bounded, the higher order terms disappear at least as fast as $n \times 1/n^2$, leaving just the average of r_n, which we will call r. Thus, in the large n limit, $\ln(Q) = r$, so $Q = \exp(r)$, which means the result depends only on the average value of r_n, and not on the variations. This argument applies for Equation (A.25) with the matrices as well. We need to be careful because the elements of this matrix are not bounded, but because the matrices are so close to diagonal, with the size of elements growing farther away from the upper-left, as long as we restrict ourselves to the upper-left corner (small electron numbers), this argument works. In that case, in the continuum limit both $g$and $l$go to zero as $1/n$, and they take the place of r_n/n in the above argument. Variations will only affect the large-number side of the histogram, and are the same order of correction as the other $1/n^2$ terms we have dropped. We will typically look at histograms on a logarithmic scale, where all the corrections will be suppressed by a factor of $1/n$ compared to the terms we keep, with n being on the order of several hundred.

and associated eigenvectors as follows:

$$\begin{aligned} \widetilde{\mathfrak{g}}\,\overrightarrow{\mathfrak{e}}_i &= \lambda_i \overrightarrow{\mathfrak{e}}_i && \text{(A.26)} \\ \lambda_i &= -\left(ig + l\right) && \text{(A.27)} \\ i &\in \{0, 1, 2, 3, \ldots\} \end{aligned}$$

Furthermore, we can expand the matrix multiplication in Equation (A.26), and exploit the fact the there are only two non-zero values in each row of $\widetilde{\mathfrak{g}}$ to come up with a simple recursion relation for the elements of $\overrightarrow{\mathfrak{e}}_i$. The j^{th} element of Equation (A.26) is:

$$\begin{aligned} \lambda_i \mathfrak{e}_{i,j} &= \left[\widetilde{\mathfrak{g}}\,\overrightarrow{\mathfrak{e}}_i\right]_j \\ &= \begin{cases} -l\mathfrak{e}_{i,j} & j = 0 \\ \left(\left(j-1\right)g + l\right)\mathfrak{e}_{i,j-1} - \left(jg + l\right)\mathfrak{e}_{i,j} & j \in \{1, 2, 3, \ldots\} \end{cases} && \text{(A.28)} \\ \lambda_i &= -\left(ig + l\right) && \text{(A.29)} \end{aligned}$$

The two cases are simply whether or not there is an $(j-1)^{\text{th}}$ column in the matrix. There is if $j > 0$, and there is not for the first ($j = 0$) column. The "recursion" for $j = 0$ is the same as the $j > 0$ case, but without the $j-1$ part: $\lambda_i \mathfrak{e}_{i,j} = -\left(jg + l\right)\mathfrak{e}_{i,j}$. When we substitute $j = 0$, this reduces to the $j = 0$ case in Equation (A.28).

At this point, we will briefly look at the solutions to two limiting cases of Equation (A.28). The simpler case is $g = 0$. From Equation (A.27), $\lambda_i = -l$, and the recursion reduces to $\mathfrak{e}_{i,0} = \mathfrak{e}_{i,0}$, $\mathfrak{e}_{i,j} = -\mathfrak{e}_{i,j-1} + \mathfrak{e}_{i,j}$. The first condition is trivial, and the second requires $\mathfrak{e}_{i,j-1} = 0$. This means the only eigenvector of $\widetilde{\mathfrak{g}}$ with $g = 0$ is the null vector. If we had not specified $\lambda_i = -l$, then we would be able to solve the recurrence relation with $\mathfrak{e}_{\lambda,j} = \left(l + \lambda\right)^{-j}$, except the $j = 0$ case requires either $\mathfrak{e}_{\lambda,0} = 0$ or $l + \lambda = 0$, which invalidates this solution. The reason for this is simple: In the $g = 0$ case, the leakage effectively shifts part of the probability for having j electrons to the $j + 1$-electron bin, by spontaneously adding an electron with probability l. Thus, each histogram bin loses some probability to the next bin up, but simultaneously gains some probability from the previous bin. You can create a steady-state of this process if you have an infinitely long chain, but if your chain has a starting point (no fewer than 0 electrons), you have one element that loses probability but does *not* gain electrons. This break in symmetry means that the first element (probability for 0 electrons) decays to 0, and once it gets close to 0, the probability for having 1 electron starts to decay (since it no longer has any gain), eventually resulting in a steady state of just the 0 vector (all probability gets spread out over an infinite number of elements). Since the $g = 0$ case has no steady-state solutions (eigenvectors), we can expect our final solution to be singular in the $g = 0$, $l \neq 0$ limit. It will turn out that the equations have a removable singularity at this point, meaning that while the formula is technically undefined, it will have a well-defined limit as $g \to 0$ ($G \to 1$). A similar effect can be seen in Equation (A.12) and Equation (A.13), where if we let $g \to 0$, which means $G \to 1$ so that $L_G = L/\ln(G)$ diverges, making the formulas singular. The singularities are removable, however, as discussed in that section.

The second special case of the recurrence relation in Equation (A.28) that is worth looking at is the $l = 0$ case. In this case, the recursion relation in Equation (A.28) becomes $-i\mathfrak{e}_{i,0} = 0$ and $-i\mathfrak{e}_{i,j} = (j-1)\,\mathfrak{e}_{i,j-1} - j\mathfrak{e}_{i,j}$ (the g factors all cancel). The second relation can be re-written as $(j-i)\,\mathfrak{e}_{i,j} = (j-1)\,\mathfrak{e}_{i,j-1}$. This relation is solvable. For the first eigenvector ($i = 0$), the first relation allows for the first element $j = 0$ to be nonzero. The second relation forces the second element to be zero times the first element, and every subsequent element to be a multiple of the previous element (which is always zero). Thus, the $i = 0$ eigenvector is the first column of the identity matrix. For $i > 0$, the first relation requires that $\mathfrak{e}_{i,0} = 0$. Again, the second requires that each be a multiple of that, until $j = i$, at which point the relation gives $0\mathfrak{e}_{i,j=i} = (j-1)\,0$, which allows this element to be non-zero. Every subsequent element ($j > i$) will be a ratio of integers times the previous element ($(j-1)\,/\,(j-i) \times \mathfrak{e}_{i,j-1}$), so if we pick 0 for $\mathfrak{e}_{i,j=i}$, we will just have a vector of zeros. Since we can scale the eigenvector by an arbitrary constant and still have an eigenvector, we can arbitrarily pick this nonzero value to be 1. The next element will be $i/1$ times this, and the next will be $(i+1)\,/2$ times that element, and then

$(i+2)/3$, etc. This ratio-of-integer progression produces a familiar pattern:

$$\begin{aligned}\mathfrak{e}_{i,j} &= \prod_{a=1}^{j-i} \frac{i+a-1}{a} = \frac{(j-1)!}{(i-1)!\,(j-i)!} \\ &=: \binom{j-1}{i-1} \text{ for } l=0 \qquad \text{(A.30)} \\ \begin{bmatrix} \overrightarrow{\mathfrak{e}}_0 & \overrightarrow{\mathfrak{e}}_1 & \overrightarrow{\mathfrak{e}}_2 & \cdots \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0 & 0 & \\ 0 & 1 & 0 & 0 & \\ 0 & 1 & 1 & 0 & \cdots \\ 0 & 1 & 2 & 1 & \\ & & \vdots & & \ddots \end{bmatrix} \\ \binom{-1}{-1} &:= 1 \end{aligned}$$

where we define the product to be 1 if the upper limit is $a = j - i = 0$ and 0 if the upper limit is $a = j - i < 0$. Here, we see the matrix of eigenvectors reproduces Pascal's Triangle.

We can use this general method to solve the general recursion in Equation (A.28). We write the recursion as:

$$\begin{aligned}(j-i)\,g\mathfrak{e}_{i,j} &= ((j-1)\,g+l)\,\mathfrak{e}_{i,j-1} \\ \mathfrak{e}_{i,-1} &:= 0\end{aligned}$$

Defining $\mathfrak{e}_{i,-1} = 0$ is just a trick to combine the two recursion relations into one statement. Starting with that point ($j = -1$), we use the recursion relation to find $\mathfrak{e}_{i,j<i} = 0$. For $j = i$, though, the recursion relation gives $0 \times \mathfrak{e}_{i,j=i} = 0$, which lets us pick $\mathfrak{e}_{i,j=i} = 1$. Then, we use the recursion relation again to show that each subsequent element of the eigenvector is $((j-1)\,g+l)\,/\,(j-i)\,g$ times the previous element. While this is no longer a ratio of integers, it is easily represented in the same product form used in Equation (A.30):

$$\mathfrak{e}_{i,j} = \prod_{a=1}^{j-i} \frac{(i+a-1)\,g+l}{ag} \qquad \text{(A.31)}$$

The $l = 0$ case shown in Equation (A.30) could be written using factorials, taking advantage of the fact that we can write the product of sequential integers as the ratio of factorials. The denominator in Equation (A.31) is a factorial, but the numerator is not. We can make the numerator a product of integer-separated numbers by factoring g out of both the numerator and the denominator. That leaves us with the ratio l/g. In the continuous limit, $l = L/n$, and $g = \ln(G)/n$, so the ratio $l/g = L/\ln(G)$, which is what we previously defined L_G to be. With this, Equation (A.31) becomes:

$$\mathfrak{e}_{i,j} = \prod_{a=1}^{j-i} \frac{(i+a-1)+L_G}{a} \tag{A.32}$$

Now the top is a product of integer-separated numbers, but they are not themselves integers. We also note the appearance of L_G, which introduces the same $g = 0$ singularity we had in Equation (A.12) and Equation (A.13), and which meshes nicely with our difficulty finding eigenvectors for the $g = 0$ case of Equation (A.28).

There is a very common function that is often considered to be a continuous version of a factorial, called the Gamma function:

$$\Gamma(n) := \int_0^\infty x^{n-1} e^{-x}\,\mathrm{d}x$$

The Gamma function is very commonly tabulated and routines for evaluating it are included in most computational packages, and it will allow us to compute the values in Equation (A.31) more easily than the product. Although the Gamma function has many well-known properties we will only use the fact that it is defined for all positive reals and this property (shown using integration by parts):[2]

$$\Gamma(n) = \int_0^\infty x^{n-1} e^{-x}\,\mathrm{d}x$$

use integration by parts with $u = x^{n-1}$, $\mathrm{d}v = e^{-x}\,\mathrm{d}x$:

$$= -\,(n-1)\,x^{n-2}e^{-x}\Big|_0^\infty + (n-1)\int_0^\infty x^{n-2}e^{-x}\,\mathrm{d}x$$

assume $n > 1$, so that the boundary term evaluates to $0 - 0 = 0$:

$$= (n-1)\,\Gamma(n-1) \qquad \text{for } n > 1$$

[2] The Gamma function can actually be defined for all complex numbers except zero and the negative real integers.

It is easy enough to directly to the integral to show $\Gamma(1) = 1$, which, combined with the above property, can be used to show that $\Gamma(n) = (n-1)!$, for positive integer n. However, this relation is not restricted to integer arguments, and we can generalize it to show:

$$\begin{aligned}\Gamma(x) &= (x-1)\,\Gamma(x-1)\\ &= (x-1)\,(x-2)\,\Gamma(x-2)\\ &= (x-1)\,(x-2)\,(x-3)\,\Gamma(x-3)\\ &= \dots\end{aligned}$$

which can be turned into

$$\frac{\Gamma(x)}{\Gamma(x-n)} = \prod_{a=1}^{n}(x-n-a-1)$$

for a positive integer n, as long as $x - n > 0$. This is exactly the sort of product that appears in the numerator of our Equation (A.32), which we can now write as:

$$\mathfrak{e}_{i,j} = \prod_{a=1}^{j-i}\frac{(i+a-1)+L_G}{a} = \frac{\Gamma(j+L_G)}{\Gamma(i+L_G)\Gamma(j-i+1)} \tag{A.33}$$

We took the liberty of using $(j-i)! = \Gamma(j-i+1)$. We should also stress that the Gamma function has not been produced by these functions (but it almost has). We are using it only because it is a common function that happens to reduce a product of integer-separated values into a ratio of two functions. This is not enough to uniquely specify the Gamma function.[3]

Now that we have a formula for the eigenvalues and eigenvectors of $\widetilde{\mathfrak{g}}$, we can continue with our quest to find an analytic formula for Equation (A.25). The next step is to decompose arbitrary initial histogram into these eigenvectors. We will then compute the effect of $\left(\widetilde{\mathbb{I}} + g\widetilde{g} + \widetilde{ll}\right)^n$ on the eigenvectors, which is very easy, and then sum the eigenvectors together again to find the resulting output

[3] As a simple example, $2\Gamma(n)$ also satisfies these relations. However, this *almost* specifies the Gamma function. The Bohr-Mollerup theorem proves that this property ($\Gamma(x) = (x-1)\,\Gamma(x-1)$), along with specifying one point ($\Gamma(1) = 1$) and requiring that the function be logarithmically convex, *does* uniquely specify the Gamma function for positive real numbers.

histogram. Since we represent histograms as probabilities of having 0 electrons as the first element, and probabilities of having 1 electron as the second element, etc., we choose to compute the effect of the EM stage on an initial histogram composed of exactly j electrons. This is represented by the j^{th} column of the identity matrix, which we will denote as:

$$\mathbb{I}_k := \begin{bmatrix} \left.\begin{matrix} 0 \\ 0 \\ \vdots \\ 0 \end{matrix}\right\} k\ 0\text{'s} \\ 1 \\ 0 \\ 0 \\ \vdots \end{bmatrix} \tag{A.34}$$

Once we know the effect of the EM stage on $\mathbb{I}_k$, we can easily compute the effect of the EM stage on an arbitrary histogram.

We note that the eigenvectors of $\widetilde{\mathfrak{g}}$ given by Equation (A.31) form a lower-diagonal matrix, which helps a great deal in finding an eigenvector-decomposition of $\mathbb{I}_k$. Assume

$$\mathbb{I}_k = \sum_{i=0}^{\infty} c_{k,i} \overrightarrow{\mathfrak{e}}_i$$

We can get some insight by viewing one such decomposition as follows (using $k = 3$):

$$\mathbb{I}_3 = c_{3,0}\overrightarrow{\mathfrak{e}}_0 + c_{3,1}\overrightarrow{\mathfrak{e}}_1 + c_{3,2}\overrightarrow{\mathfrak{e}}_2 + c_{3,3}\overrightarrow{\mathfrak{e}}_3 + c_{3,4}\overrightarrow{\mathfrak{e}}_4 + \cdots$$

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ \vdots \end{bmatrix} = c_{3,0} \begin{bmatrix} 1 \\ * \\ * \\ * \\ * \\ \vdots \end{bmatrix} + c_{3,1} \begin{bmatrix} 0 \\ 1 \\ * \\ * \\ * \\ \vdots \end{bmatrix} + c_{3,2} \begin{bmatrix} 0 \\ 0 \\ 1 \\ * \\ * \\ \vdots \end{bmatrix} + c_{3,3} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ * \\ \vdots \end{bmatrix} + c_{3,4} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \\ \vdots \end{bmatrix} + \cdots$$

where we have used '$*$' to denote possibly non-zero values. By looking across the top row, we see the only non-zero element is from $\overrightarrow{\mathfrak{e}}_0$. Since the top element of $\mathbb{I}_3$

is 0, we therefore know the coefficient of $\overrightarrow{\mathfrak{e}}_0$, $c_{3,0}$ must be 0. Looking across the second row, we see the only non-zero values are from $\overrightarrow{\mathfrak{e}}_0$ and $\overrightarrow{\mathfrak{e}}_1$. The coefficient of $\overrightarrow{\mathfrak{e}}_0$ is 0, so $\overrightarrow{\mathfrak{e}}_1$ contributes the only non-zero value to the second row. Since the second value of $\mathbb{I}_3$ is 0, we know $c_{3,1} = 0$. Likewise, looking at the third row, only $\overrightarrow{\mathfrak{e}}_{i\leq 2}$ contribute non-zero values. However, since the coefficients for $\overrightarrow{\mathfrak{e}}_{i<2}$ are 0, only $\overrightarrow{\mathfrak{e}}_2$ contributes to this row. Since $\mathbb{I}_3$ is zero in this row, we know $c_{3,2} = 0$. This argument continues, with $c_{3,i} = 0$, until we reach row 3, where we get $c_{3,3} = 1$, because $\mathbb{I}_3$ has a 1 in that row instead of a 0. More generally, if we are finding the decomposition of $\mathbb{I}_k$, then this argument that shows $c_{j,i} = 0$ will work for $i = 0$ up to (but not including) $i = k$, at which point we will get $c_{k,i=k} = 1$.

For $i > k$, the process becomes more difficult, because we no longer have $c_{k,a} = 0$ for $a < i$, but a similar process works. To compute $c_{k,i}$, we assume we have worked our way up from $i = 0$ to the current value, so that we know $c_{k,a<i}$. We look at the i^{th} row. The only eigenvectors with non-zero elements in the i^{th} row are $\overrightarrow{\mathfrak{e}}_{a\leq i}$. We know the coefficients of all of those vectors except for $\overrightarrow{\mathfrak{e}}_{a=i}$, and we know what it needs to sum to, so we can easily compute what the coefficient of $\overrightarrow{\mathfrak{e}}_{a=i}$ is.

We found it easier to work with the $l = 0$ case, where the eigenvector elements were binomial coefficients. After working out several of the coefficients using the above iterative method, we recognized a pattern:

$$c_{k,i} = (-1)^{i-k} \binom{k-1}{i-1} \text{ for } l = 0$$

We note that we have defined $\binom{-1}{-1} = 1$, and $\binom{i}{k>i} = 0$, so that this version encompasses $c_{k,i<k} = 0$. In particular, we noted this relation:

$$c_{k,i} = (-1)^{i-k}\, \mathfrak{e}_{k,i} \text{ for } l = 0 \tag{A.35}$$

We were able to prove this relation works using a rather handy version of the Binomial Theorem. In general, the Binomial Theorem states:

$$(a+b)^n = \sum_{i=0}^{n} \binom{n}{i} a^i b^{n-i} \tag{A.36}$$

This can be proved relatively easily with an induction argument, or with the following combinatoric argument: If we write out $(a + b)$ n times, then we can generate terms of the product by picking either a or b in each factor. The product is the sum of all such factors (this is basically a generalized distributive property of how addition distributes through multiplication). Once like terms are collected, the coefficient of the a^i term will be the number of ways we can pick i of the n factors (and we pick a from those i terms and b from the remaining $n - i$ terms), which is $\binom{n}{i}$.[4] The particular version of the Binomial Theorem we will use is:

$$(1 + a)^n = \sum_{i=0}^{n} \binom{n}{i} a^i \tag{A.37}$$

This is Equation (A.36) with $b = 1$, and it states that a geometric series (a^i) with binomial coefficient weights can be summed. We will use this in proving our final formula for eigenvector decomposition of identity-matrix columns.

It turns out that our final solution for the eigenvalue decomposition of a column of the identity matrix is the same of one of our forms for the $l = 0$ case, Equation (A.35):

$$\begin{aligned}
\mathbb{I}_k &= \sum_{i=0}^{\infty} c_{k,i} \overrightarrow{\mathfrak{e}}_i \\
c_{k,i} &= (-1)^{i-k} \, \mathfrak{e}_{k,i} \\
&= \begin{cases} 0 & i < k \\ 1 & i = k \\ (-1)^{i-k} \left\{ \begin{array}{c} \displaystyle\prod_{a=1}^{i-k} \frac{(k + a - 1)\, g + l}{ag} \\ \text{or} \\ \displaystyle\prod_{a=1}^{i-k} \frac{(k + a - 1) + L_G}{a} \end{array} \right\} & i > k \end{cases}
\end{aligned} \tag{A.38}$$

[4] Another way to demonstrate this theorem is to use the fact that each element in Pascal's Triangle is the sum of the element above it with the element above and to the left of it: $\binom{n}{i} = \binom{n-1}{i} + \binom{n-1}{i-1}$. If we weight each element (row n, column i) of Pascal's Triangle with the factor $a^i b^{n-i}$, then the element above is $a^i b^{n-i-1}\binom{n-1}{i}$, and the element above and left is $a^{i-1} b^{n-i} \binom{n-i}{i-1}$ (n and i are both decreased, so $n - i$ is unchanged). Our new sum rule is that each element is b times the element above plus a times the element above and left. If we sum an entire row, that works out to be a times the sum of the row above plus b times the sum of the row above, or $(a + b)$ times the sum of the row above. Given that the sum of the first row is 1, the sum of the n^{th} row must therefore be $(a + b)^n$.

We typically define the product to be 1 if $i - k = 0$ and 0 if $i - k < 0$, but we have explicitly written those cases out separately here.

We will now prove Equation (A.38). Start with the sum:

$$\overrightarrow{S}_k := \sum_{i=0}^{\infty} c_{k,i} \overrightarrow{\mathfrak{e}}_i$$

We wish to prove that this sum is $\mathbb{I}_k$ when we use $c_{k,i}$ as given by Equation (A.38). $c_{k,i} = 0$ for $i < k$, which means we can restrict the sum to $i = k$ to ∞. Furthermore, we will now restrict ourselves to one row (row j):

$$\overrightarrow{S}_k \left(\text{row } j\right) = \sum_{i=k}^{\infty} c_{k,i} \mathfrak{e}_{i,j}$$

Now, much like in the iterative method we used to compute these coefficients, we note that $\mathfrak{e}_{i,j}$ is non-zero only when $j \geq i$. Therefore, the sum is only non-zero when it contains some terms with $i < j$, so we can set the upper limit of the sum to j:

$$\overrightarrow{S}_k \left(\text{row } j\right) = \begin{cases} 0 & j < k \\ \displaystyle\sum_{i=k}^{j} c_{k,i} \mathfrak{e}_{i,j} & j \geq k \end{cases} \tag{A.39}$$

Now, we substitute the values for $c_{k,i}$ and $\mathfrak{e}_{i,j}$ (from Equation (A.31) and Equation (A.38)) and start manipulating the $j \geq k$ case:

$$\sum_{i=k}^{j} c_{k,i} \mathfrak{e}_{i,j} = \sum_{i=k}^{j} \left(\left(-1\right)^{i-k} \prod_{a=1}^{i-k} \frac{\left(k + a - 1\right) g + l}{ag} \right) \left(\prod_{a=1}^{j-i} \frac{\left(i + a - 1\right) g + l}{ag} \right)$$

In the products, do the numerators and denominators separately. The denominators are simply factorials and powers:

$$= \sum_{i=k}^{j} \left(-1\right)^{i-k} \frac{\prod_{a=1}^{i-k} \left[\left(k + a - 1\right) g + l\right]}{(i-k)! g^{i-k}} \frac{\prod_{a=1}^{j-i} \left[\left(i + a - 1\right) g + l\right]}{(j-i)! g^{j-i}}$$

Replace a with $a-k+1$ in the first product and $a-i+1$ in the second. The first product is then from $a-k+1=1$ to $a-k+1=i-k$, which can also be written as $a=k$ to $i-1$. The second product is similar:

$$= \sum_{i=k}^{j} (-1)^{i-k} \frac{\prod_{a=k}^{i-1} [ag+l]}{(i-k)! g^{i-k}} \frac{\prod_{a=i}^{j-1} [ag+l]}{(j-i)! g^{j-i}}$$

Notice that the first product covers a from k to $i-1$, and the second picks up where that one left off with i and goes to $j-1$, so we can combine them. The combination is independent of i, and so we can pull it out of the sum. Likewise, we can combine the g factors in the denominators, and pull them out. Finally, we multiply the numerator and denominator by $(j-k)!$:

$$= \frac{\prod_{a=k}^{j-1} [ag+l]}{(j-k)! g^{j-k}} \sum_{i=k}^{j} (-1)^{i-k} \frac{(j-k)!}{(i-k)!\,(j-i)!}$$

$$= \frac{\prod_{a=k}^{j-1} [ag+l]}{(j-k)! g^{j-k}} \sum_{i=k}^{j} (-1)^{i-k} \binom{j-k}{i-k}$$

Replace i with $i+k$ in the sum, and re-write the limits in terms of the new i. We can then apply A.37 to remove the sum:

$$= \frac{\prod_{a=k}^{j-1} [ag+l]}{(j-k)! g^{j-k}} \sum_{i=0}^{j-k} (-1)^{i} \binom{j-k}{i}$$

$$= \frac{\prod_{a=k}^{j-1} [ag+l]}{(j-k)! g^{j-k}} (1-1)^{j-k}$$

The $j = k$ case of the sum consisted of only one term (1), and so is 1. The new form of the sum makes it obvious that the sum is 0 for $j \neq k$:

$$= \frac{\prod_{a=k}^{j-1} [ag + l]}{(j-k)!g^{j-k}} \begin{cases} 0 & j \neq k \\ 1 & j = k \end{cases}$$

We now recall that we were computing

$$\sum_{i=k}^{j} c_{k,i} \mathfrak{e}_{i,j}$$

for $j \geq k$, and we have now shown that it is 0 for $j > k$. We can either use our above result to compute the $j = k$ case, or directly from the definitions, using $c_{k,i=k} = 1$ and $\overrightarrow{\mathfrak{e}}_{i} j = i = 1$:

$$\sum_{i=k}^{k} c_{k,i} \mathfrak{e}_{i,j} = c_{k,k} \mathfrak{e}_{k,k} = (1)\,(1) = 1$$

Combining this with Equation (A.39), we end up with:

$$\overrightarrow{S}_k \,(\text{row } j) = \begin{cases} 0 & j < k \\ 1 & j = k \\ 0 & j > k \end{cases}$$

This is precisely the j^{th} row of $\mathbb{I}_k$. Therefore,

$$\overrightarrow{S}_k = \mathbb{I}_k$$

and this proves that the decomposition given in Equation (A.38) works.

Now that we can decompose $\mathbb{I}_k$ into eigenvectors of $\widetilde{\mathfrak{g}}$, we can rather easily use this with Equation (A.25) to find out what the EM stage does to a known number of electrons coming in. Assuming $\mathbb{I}_k$ is the input histogram for the EM stage, and that we are in the continuous limit, then we start from Equation (A.25) and get:

$$\mathcal{H}\{N_n\} = \left(\widetilde{\mathbb{I}} + \widetilde{\mathfrak{g}}\right)^n \mathbb{I}_k$$

Apply Equation (A.38):

$$= \left(\widetilde{\mathbb{I}} + \widetilde{\mathfrak{g}}\right)^n \sum_{i=0}^{\infty} c_{k,i} \overrightarrow{\mathfrak{e}}_i$$

Equation (A.26) and Equation (A.27) imply that $\left(\widetilde{\mathbb{I}} + \widetilde{\mathfrak{g}}\right) \overrightarrow{\mathfrak{e}}_i = (1 - ig - l) \overrightarrow{\mathfrak{e}}_i$, which, repeated n times, yields:

$$= \sum_{i=0}^{\infty} c_{k,i} \left(1 - ig - l\right)^n \overrightarrow{\mathfrak{e}}_i$$

Paralleling the proof of Equation (A.38), we now look at the j^{th} element of this vector, and introduce the values of $c_{k,i}$ and $\mathfrak{e}_{i,j}$ from Equation (A.38) and Equation (A.31), respectively. The j^{th} element of $\mathcal{H}\{N_n\}$ is the probability of having j electrons after the EM stage, which we have been, and will continue, denoting as $\mathcal{H}\{N_n\}(j)$:

$$\begin{aligned}\mathcal{H}\{N_n\}(j) &= \sum_{i=0}^{\infty} c_{k,i} \left(1 - ig - l\right)^n \mathfrak{e}_{i,j} \\ &= \sum_{i=k}^{j} \left(1 - ig - l\right)^n \times \\ &\quad \left((-1)^{i-k} \prod_{a=1}^{i-k} \frac{(k+a-1)\,g + l}{ag}\right) \left(\prod_{a=1}^{j-i} \frac{(i+a-1)\,g + l}{ag}\right)\end{aligned}$$

We have dropped the $i < k$ terms of the sum, because $c_{k,i} = 0$ for those cases. Likewise, because $\mathfrak{e}_{i,j} = 0$ for $i > j$, we have eliminated the $i > j$ terms of the sum. Because all terms are 0 when $j < k$, we just remember that this sum is 0 for $j < k$, and just work on the $j \geq k$ case. Again, the products from $a = 1$ to 0 represent quantities that are actually 1, and we choose to just define the $\prod$ notation to mean this rather than continually writing out separate cases (those cases are easy to check separately, as most of the factors become 1 and the sums have only one term). Exactly as we did in proving Equation (A.38), we can manipulate the products into a single product that is independent of i, and turn the terms that are dependent on i into a combinatorial factor, up to the point where we used the

Binomial Theorem before:

$$\begin{aligned}\mathcal{H}\{N_n\}(j) &= \sum_{i=k}^{j} (1-ig-l)^n\,(-1)^{i-k}\,\frac{\prod\limits_{a=k}^{i-1}[ag+l]}{(i-k)!g^{i-k}}\,\frac{\prod\limits_{a=i}^{j-1}[ag+l]}{(j-i)!g^{j-i}}\\ &= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}}\sum_{i=k}^{j}(1-ig-l)^n\,(-1)^{i-k}\,\frac{(j-k)!}{(i-k)!\,(j-i)!}\\ &= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}}\sum_{i=k}^{j}(1-ig-l)^n\,(-1)^{i-k}\binom{j-k}{i-k}\end{aligned}$$

This is still for the $j \geq k$ case. Right now, we have defined the sum to be 0 when $j < k$, but once we explicitly calculate the sum, that will be lost. We will remember this relation by defining the $\prod$ notation to be 1 when the upper limit is one less than the lower limit ($j = k$), and 0 when the upper limit is even less that.

This is the point where we used the Binomial Theorem before. The only catch is we now have $(1-ig-l)^n$ inside the sum. Using the same arguments we used when we kept only first-order terms in Equation (A.25), we can again expand this to lowest order, replace g with the continuous limit from Equation (A.11), and replace l with its definition in terms of overall leakage, Equation (A.8). This then becomes a limit of the form $(1+x/n)^n \to e^x$. However, we can also simplify this in a different way, reminiscent of a common trick for using l'Hôpital's rule on 1^∞ limits: Take the logarithm. This yields:

$$\ln((1-ig-l)^n) = n\ln(1-ig-l)$$

Since both g and l go to 0 in the large n limit, and we are trying to compute the large n limit of this expression, we can assume that n is large enough that $ig+l$ is less than 1. This lets us replace the logarithm with a power-series expansion:

$$= n\left\{(-ig-l) + \mathcal{O}\big(g^2, l^2, gl\big)\right\}$$

Now, if we take the large n limit, we can use Equation (A.11) and Equation (A.8) to write:

$$
\begin{aligned}
&= -ing - nl + n\mathcal{O}\big(g^2, l^2, gl\big) \\
&= -i\ln(G) - L
\end{aligned}
$$

We dropped the higher order terms, because they go to zero. For example, $ng^2 = ng \times g$, which is a finite quantity (ng) times a quantity which goes to zero (g), so the whole product goes to zero. This shows that $\ln((1 - ig - l)^n)$ is very closely approximated by $-i\ln(G) - L$ in the continuous limit, which means:

$$(1 - ig - l)^n = e^{-i\ln(G) - L} = G^{-i}e^{-L} \text{ as } n \to \infty$$

Inserting this back into our sum yields:

$$
\begin{aligned}
\mathcal{H}\{N_n\}(j) &= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}} \sum_{i=k}^{j} (1 - ig - l)^n \, (-1)^{i-k} \binom{j-k}{i-k} \\
&= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}} \sum_{i=k}^{j} \left(G^{-i}e^{-L}\right) (-1)^{i-k} \binom{j-k}{i-k} \\
&= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}} e^{-L}G^{-k} \sum_{i=k}^{j} G^{-(i-k)} \, (-1)^{i-k} \binom{j-k}{i-k}
\end{aligned}
$$

Replace i with $i+k$ in the sum, change the limits, and apply the Binomial Theorem, Equation (A.37):

$$
\begin{aligned}
&= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}} e^{-L}G^{-k} \sum_{i=0}^{j-k} \left(-\frac{1}{G}\right)^{i} \binom{j-k}{i} \\
&= \frac{\prod\limits_{a=k}^{j-1}[ag+l]}{(j-k)!g^{j-k}} e^{-L}G^{-k} \left(1 - \frac{1}{G}\right)^{j-k}
\end{aligned}
$$

Finally, we can get rid of g by canceling each power of g in the denominator with one factor of g in the product in the numerator. Once again, we use $l/g = (L/n)\,/\,(\ln(G)/n) = L/\ln(G) =: L_G$, and we also show the result using Gamma functions, which becomes possible because, once we cancel g, the factors in the product are separated by integers:

$$\begin{aligned}\mathcal{H}\{N_n\}(j) &= \frac{\prod_{a=k}^{j-1}[a+L_G]}{(j-k)!}e^{-L}G^{-k}\left(1-\frac{1}{G}\right)^{j-k} \\ &= \frac{\Gamma(j+L_G)}{\Gamma(k+L_G)\Gamma(j-k+1)}e^{-L}G^{-k}\left(1-\frac{1}{G}\right)^{j-k} \qquad \text{(A.40)}\end{aligned}$$

This is the histogram that comes out of the EM stage when a distinct number of electrons (k) enter the EM stage. If the incoming histogram is not well-defined, then we can simply write it as a weighted sum of well-defined inputs:

$$\mathcal{H}\{N_0\} = \mathcal{H}\{N_0\}(0)\mathbb{I}_0 + \mathcal{H}\{N_0\}(1)\mathbb{I}_1 + \mathcal{H}\{N_0\}(2)\mathbb{I}_2 + \cdots$$

Since the EM stage transforms histograms in a linear fashion (in this model), the output histogram is the same sum over these well-defined inputs. For completeness, we also add in the convolution with the readout noise:

$$\begin{aligned}\mathcal{H}\{\text{final}\} &= \mathcal{H}\{N_n\} * \mathcal{H}\{\text{readout noise}\} \\ \mathcal{H}\{N_n\}(j) &= \sum_{k=0}^{j}\frac{\Gamma(j+L_G)}{\Gamma(k+L_G)\Gamma(j-k+1)}e^{-L}G^{-k}\left(1-\frac{1}{G}\right)^{j-k}\mathcal{H}\{N_0\}(k) \qquad \text{(A.41)} \\ \mathcal{H}\{N_0\} &= \mathcal{H}\{\text{signal}\} * \mathcal{H}\{\text{CIC}\} \\ n &\to \infty\end{aligned}$$

where the readout noise function is a normalized Gaussian with center m_r and standard deviation σ_{r}, and the CIC noise function is a Poisson distribution with mean σ_{CIC}. As discussed earlier, thermal noise also has a Poissonian distribution, and the convolution of the two is another Poissonian distribution. Thus, σ_{CIC} actually refers to the mean and variance of the combined distribution (it is actually the sum of the means or variances of the actual CIC and thermal noises). We note that we can eliminate either L or L_G in favor of the other by using either $L_G = L/\ln(G)$ or $e^{-L} = G^{-L_G}$.

The sum in Equation (A.41) terminates at $k = j$ because $\mathcal{H}\{N_n\}(j)$ is 0 when $k > j$ (this model does not allow for loss of electrons). This $k \leq j$ was lost when we changed from the $\prod$ notation to the Gamma functions (except the $\Gamma(j - k + 1)$ becomes undefined when $k > j$), but in the $\prod$ notation, that was where we defined the product to be 1 when the upper limit was one less than the lower limit, and 0 when the upper limit was even less. We should mention that in the $L \to 0$ (no leaking) case, there is an apparent singularity when $k = 0$, in that there is an $\Gamma(0)$ in the denominator. Looking back at the $\prod$ version, we can see that, for $j > 0$, the product really is 0 (the first factor is $k = 0$), so the $\Gamma(0) = \infty$ in the denominator is correct. For $j = 0$, the product has limits of 0 to -1, which is one of the cases that we have defined it to be one. It is as though the $\Gamma(k = 0)$ in the denominator cancels the $\Gamma(j = 0)$ in the numerator.

Once again, we see this apparent discontinuity in the $G \to 1$ limit, as $L_G \to \infty$ in this limit. In this limit, if a known number k of electrons enter the EM stage, we can write the histogram for the output of the EM stage as follows:

$$\begin{aligned}
\mathcal{H}\{N_n\}(j) &= \frac{\prod_{a=k}^{j-1} [a + L_G]}{(j-k)!} e^{-L} G^{-k} \left(1 - \frac{1}{G}\right)^{j-k} \\
&= \frac{\prod_{a=k}^{j-1} \left[(a + L_G)\left(1 - \frac{1}{G}\right)\right]}{(j-k)!} e^{-L} G^{-k} \\
&= \frac{\prod_{a=k}^{j-1} \left[(a + L_G)\frac{G-1}{G}\right]}{(j-k)!} e^{-L} G^{-k}
\end{aligned}$$

As mentioned in showing the $G \to 1$ limit of Equation (A.12) and Equation (A.13) were well-defined, $(G-1)\,L_G \to L$ in the $G \to 1$ limit:

$$= \frac{\displaystyle\prod_{a=k}^{j-1} \frac{a\,(G-1)+L}{G}}{(j-k)!} e^{-L} G^{-k}$$

Now we take $G \to 1$ wherever G appears:

$$= \frac{\displaystyle\prod_{a=k}^{j-1} L}{(j-k)!} e^{-L}$$
$$= \frac{L^{j-k}}{(j-k)!} e^{-L}$$

We note that we have arrived at a Poissonian distribution that starts at $j = k$ (and is 0 for $j < k$), with a mean and variance of L. Since we set up the leakage to be a Poissonian distribution with a mean of L in the no-gain limit, this is exactly what we should get.

We would also like to note the $L \to 0$ limit, where there is no leakage. In this limit, if you have k electrons entering the EM stage, the probability of having j electrons after the EM stage is the same as the probability of gaining $j-k$ electrons. Equation (A.41) gives this probability in the $L \to 0$ limit:

$$\mathcal{H}\{N_{n\to\infty}\}(j) = \binom{j-1}{k-1} \left(\frac{1}{G}\right)^{k} \left(1-\frac{1}{G}\right)^{j-k}. \tag{A.42}$$

This represents the distribution we would expect if a single electron were randomly multiplied as it passed through a continuous set of amplifiers, with no leakage. This is sometimes used as a model for photomultipliers. In photomultipliers, a photon excites an electron, which is accelerated through a large voltage to strike a target, exciting more electrons. This process is repeated multiple times, similar to what happens in the EM stage of an EMCCD. Some photomultipliers may be treated as a continual set of small amplification stages, exactly how we are treating the EM stage. In a photomultiplier, the time between random excitations, or dark counts, is typically at least on the order of microseconds, while the time it takes to multiply an electron is sub-nanosecond [103]. Therefore, the probability of a random excitation during electron multiplication is very small. To contrast that with an EMCCD, we find that while the probability of an extra excitation happening during a single stage of amplification is small, the probability of a random excitation happening at least once through the entire multiplication stage is non-negligible. We demonstrate that these excitations have a distinct effect in Section V.5. Therefore, we might expect result in Equation (A.41) should also pertain to certain photomultipliers in the $L \to 0$ limit, and it does match some known photomultiplier output distributions [103]. More specifically, it has been shown that any set of continual amplifications, without leakage, converges to an exponential distribution similar to Equation (A.42), for a single input electron [104]. However, we believe this treatment with leakage is original.

At this point, we would like to compare our histogram result with Equation (A.12) and Equation (A.13). In fact, we can even derive those two results from Equation (A.41). As shown in Equation (B.14) and the following discussion, the normalizations of two convolved functions multiply, and the means and variances add. We assume that $\mathcal{H}\{\text{signal}\}$ and $\mathcal{H}\{\text{CIC}\}$ are normalized, so $\mathcal{H}\{N_0\}$ is normalized. To keep with the notation used in Equation (A.12) and Equation (A.13), we will use M and V as the mean and variance of the signal, and σ_{CIC} as both the mean and variance of the CIC (which are the same). Therefore, the mean and variance entering the EM stage are:

$$\text{mean}(\mathcal{H}\{N_0\}) = M + \sigma_{\text{CIC}}$$
$$\text{var}(\mathcal{H}\{N_0\}) = V + \sigma_{\text{CIC}}$$

We next compute three moments of the formula for $\mathcal{H}\{N_n\}$:

$$\langle 1 \rangle = \sum_{j=0}^{\infty} \mathcal{H}\{N_n\}$$
$$\langle j \rangle = \sum_{j=0}^{\infty} \mathcal{H}\{N_n\} j$$
$$\left\langle j^2 \right\rangle = \sum_{j=0}^{\infty} \mathcal{H}\{N_n\} j^2$$

The first will show that $\mathcal{H}\{N_n\}$ has the same normalization as $\mathcal{H}\{N_0\}$, which is important if we are to interpret $\mathcal{H}\{N_n\}$ as either a probability distribution or a histogram. The second is the mean value after the EM stage, and will reduce to Equation (A.12). The third can be used to compute the variance, which will reduce to Equation (A.13). All three of these are of the form $\left\langle j^b \right\rangle$, with $b = \{0, 1, 2\}$, so let us look at this result, and actually substitute the value for $\mathcal{H}\{N_n\}$ in:

$$\begin{aligned}
\left\langle j^b \right\rangle &= \sum_{j=0}^{\infty} j^b \mathcal{H}\{N_n\} \\
&= \sum_{j=0}^{\infty} j^b \sum_{k=0}^{j} \frac{\Gamma(j + L_G)}{\Gamma(k + L_G)\Gamma(j - k + 1)} e^{-L} G^{-k} \left(1 - \frac{1}{G}\right)^{j-k} \mathcal{H}\{N_0\}(k)
\end{aligned}$$

The pre-EM histogram is assumed to be normalized, so that it sums to 1. For any particular k value, we could compute the moments of the post-EM histogram using the same methods we will use shortly. In particular, the sums all converge absolutely (they converge, and the terms are all non-negative). Since we are taking a collection of sums with terms that converge absolutely, weighting them by non-negative weights that sum to 1 ($\mathcal{H}\{N_0\}(k)$), and summing those, the combined sum will still converge absolutely. This allows us to re-order the sums. We need only watch the limits, which are to cover all j, k such that $j \geq k$:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=k}^{\infty} j^b \frac{\Gamma(j + L_G)}{\Gamma(k + L_G)\Gamma(j - k + 1)} \left(1 - \frac{1}{G}\right)^{j-k}$$

We now substitute $x = (1 - 1/G)$, and return to $\prod$ notation:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=k}^{\infty} j^b \left(\prod_{a=k}^{j-1} [a + L_G] \right) \frac{x^{j-k}}{(j - k)!}$$

Finally, we replace j with $j + k$ and a with $a + k$, and change limits accordingly:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} (j + k)^b \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!}$$

Here is an “and then a miracle occurs” step, but we will try to motivate it. First, we look at the j sum with $L_G = 0$, $b = 0$, and $k = 1$:

$$\sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a+1] \right) \frac{x^j}{j!} = \sum_{j=0}^{\infty} j! \frac{x^j}{j!} = \sum_{j=0}^{\infty} x^j$$

This is a geometric series, which sums to $(1-x)^{-1}$. Now, we look at the same sum with $L_G = 0$ and $b = 0$, but now with $k = 2$:

$$\sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a+2] \right) \frac{x^j}{j!} = \sum_{j=0}^{\infty} \frac{(j+1)!}{1!} \frac{x^j}{j!} = \sum_{j=0}^{\infty} (j+1)\, x^j$$

This is the derivative of the $k = 1$ sum, and so sums to the derivative of $(1-x)^{-1}$, which is $(1-x)^{-2}$. It is not hard to show that, for all $k > 1$ (with $L_G = 0$ and $b = 0$), the sum is the $(k-1)^{\text{th}}$ derivative of the $k = 1$ case divided by $(k-1)!$, which works out to be:

$$\sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a+k] \right) \frac{x^j}{j!} = \frac{1}{(k-1)!} \frac{\mathrm{d}^{k-1}}{\mathrm{d}x^{k-1}} \frac{1}{1-x} = \frac{1}{(1-x)^k}$$

For $L_G \neq 0$, the only change to the sum is that we replace k with $k + L_G$, so we might guess that the same replacement happens on the right-hand side of the above equation. In fact, that does get the correct answer. It is easy to verify that the left-hand side is the Taylor expansion of the right-hand side, even with that replacement. The trick is showing that the right-hand side is equal to its Taylor expansion. To prove this, we will use the generalized Binomial Theorem, which is proved in many advanced calculus or real analysis textbooks:[5]

$$(1+x)^k = \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [k-a] \right) \frac{x^j}{j!} \text{ for all } k \text{ and } |x| < 1 \tag{A.43}$$

This basically states that $(1+x)^n$ is equal to its Taylor expansion, even if n is not a positive integer (the product is the successive powers brought down from

[5] Alternatively, you can use a theorem from complex analysis that states that if a function is differentiable (in the complex sense) in a disk of radius R about the origin, its Taylor series converges to that function within that disk. This function, $(1-z)^{-(k+L_G)}$, is differentiable for $|z| < 1$ (the distance to the pole at $z = 1$), which is a disk of radius 1.

the exponent k with each derivative, and is assumed to be 1 for $j = 0$). When n is a positive integer, the sum terminates at $j = k$, because all successive terms are 0 (the $a = k$ factor in the product is 0), and the product can be written as $k!/(j-j)!$, which combines with the $j!$ to create a combinatorial factor, and the original Binomial Theorem Equation (A.37) is reproduced (except that theorem holds for any x). We will use this with with the replacements $k \to -k - L_G$ and $x \to -x$, which we can do because k does not need to be an integer for the generalized Binomial Theorem:

$$(1-x)^{-k-L_G} = \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [-k - L_G - a] \right) (-1)^j \, \frac{x^j}{j!}$$

Combine the $(-1)^j$ with the product:

$$= \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \tag{A.44}$$

This is exactly the sum formula we were trying to prove. Since we defined $x = 1 - 1/G$, $x < 1$ is equivalent to $G > 1$, which is acceptable (we already showed that the $G = 1$ produces a Poissonian distribution, which agrees with Equation (A.12) and Equation (A.13) in the same limit).

We were trying to compute some moments of the post-EM histogram, and we reached this point:

$$\left\langle j^b \right\rangle = \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} (j+k)^b \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!}$$

For $b = 0$, we can use Equation (A.44) for the j sum:

$$\begin{aligned}\left\langle j^0 \right\rangle &= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \\ &= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \, (1-x)^{-k-L_G}\end{aligned}$$

We originally defined $x = 1 - 1/G$, so $1 - x = 1/G$:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) G^{k+L_G}$$

Use $L_G = L/\ln(G)$, so that $G^{L_G} = e^L$:

$$= \sum_{k=0}^{\infty} \mathcal{H}\{N_0\}(k)$$

This proves:

$$\left\langle j^0 \right\rangle = \langle 1 \rangle = \sum_{j=0}^{\infty} \mathcal{H}\{N_n\}(j) = \sum_{k=0}^{\infty} \mathcal{H}\{N_0\}(k) \tag{A.45}$$

which means that normalization is preserved—if the histogram before the EM stage sums to 1, then the histogram after the EM stage sums to 1.

Next, we would like to compute the higher-order moments. We can compute $\left\langle j^{b=1} \right\rangle$ by combining the extra j factor into the product (or realizing that we get a similar factor by taking the derivative with respect to x). Here is the trick we will use:

$$(j + k + L_G) \prod_{a=0}^{j-1} [a + k + L_G] = \prod_{a=0}^{j} [a + k + L_G]$$

$(j + k + L_G)$ is the last factor of the above product. Now, pull the first factor out:

$$= (k + L_G) \prod_{a=1}^{j} [a + k + L_G]$$

Now replace k with $k + 1$ and change the limits appropriately:

$$= (k + L_G) \prod_{a=0}^{j-1} [a + k + L_G + 1]$$

With this trick, we can compute the first moment:

$$\begin{aligned}
\left\langle j^1 \right\rangle &= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} (j + k)^1 \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \\
&= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} (j + k + L_G - L_G) \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \\
&= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \left\{ \left[(j + k + L_G) \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \right] - \right. \\
&\qquad \left. \left[L_G \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \right] \right\}
\end{aligned}$$

Using the above trick:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \left\{ \left[(k + L_G) \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G + 1] \right) \frac{x^j}{j!} \right] - \right.$$
$$\left. \left[L_G \sum_{j=0}^{\infty} \left(\prod_{a=0}^{j-1} [a + k + L_G] \right) \frac{x^j}{j!} \right] \right\}$$

Use Equation (A.44) for the two sums:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \times$$
$$\left\{ \left[(k + L_G)\,(1 - x)^{-(k+L_G+1)} \right] - \left[L_G\,(1 - x)^{-(k+L_G)} \right] \right\}$$

Use $(1 - x) = 1/G$:

$$= \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \left\{ \left[(k + L_G)\, G^{k+L_G+1} \right] - \left[L_G G^{k+L_G} \right] \right\}$$

Again, $e^{L} G^{-k}$ cancels G^{k+L_G}:

$$= \sum_{k=0}^{\infty} \mathcal{H}\{N_0\}(k) \left\{ (k + L_G)\, G - L_G \right\}$$
$$= G \left(\sum_{k=0}^{\infty} \mathcal{H}\{N_0\}(k) k \right) + (G - 1)\, L_G \left(\sum_{k=0}^{\infty} \mathcal{H}\{N_0\}(k) \right)$$

If we are to interpret $\langle j \rangle$ as the mean value, then we need to assume the histograms are normalized to 1. With that assumption, the second sum above is 1, and the first is the mean value of the pre-EM histogram, which is $M + \sigma_{\mathrm{CIC}}$. Substituting those values gives:

$$\langle j \rangle = G\,(M + \sigma_{\mathrm{CIC}}) + (G - 1)\, L_G$$

This is the mean value for the post-EM histogram. The final step is to convolve with the readout stage noise, which adds m_r to the post-EM mean, which reproduces Equation (A.12).

To compute $\left\langle j^{b=2}\right\rangle$, we use a very similar trick as for the $b=1$ case:

$$\begin{aligned}(j+k+L_G)(j+k+L_G+1)\prod_{a=0}^{j-1}[a+k+L_G] &= \prod_{a=0}^{j+1}[a+k+L_G] \\ &= (k+L_G)(k+L_G+1)\prod_{a=0}^{j-1}[a+k+L_G+2]\end{aligned}$$

We now have:

$$\left\langle j^2\right\rangle = \sum_{k=0}^{\infty} e^{-L} G^{-k} \mathcal{H}\{N_0\}(k) \sum_{j=0}^{\infty} (j+k)^2 \left(\prod_{a=0}^{j-1}[a+k+L_G]\right)\frac{x^j}{j!}$$

We next use:

$$\begin{aligned}(j+k)^2 &= ((j+k+L_G)-L_G)((j+k+L_G+1)-L_G-1) \\ &= (j+k+L_G)(j+k+L_G+1)-(j+k+L_G)(2L_G+1)+L_G^2\end{aligned}$$

The first term will use the trick given above. The second term will use the trick we used for $b=1$. The third term will just multiply the sum. If we substitute this into our formula for $\langle j^2\rangle$, use the appropriate tricks, and perform the sums, we get:

$$\begin{aligned}\left\langle j^2\right\rangle = \sum_{k=0}^{\infty} e^{-L}G^{-k}\mathcal{H}\{N_0\}(k)\Big\{&(k+L_G)(k+L_G+1)(1-x)^{-(k+L_G+2)} - \\ &(2L_G+1)(k+L_G+1)(1-x)^{-(k+L_G+1)} + \\ &L_G^2(1-x)^{-(k+L_G)}\Big\}\end{aligned}$$

Once again, use $1-x=1/G$, and cancel $e^{-L}G^{-k}$ with G^{k+L_G}:

$$\begin{aligned}&= \sum_{k=0}^{\infty}\mathcal{H}\{N_0\}(k)\big\{(k+L_G)(k+L_G+1)G^2- \\ &\qquad\qquad (2L_G+1)(k+L_G+1)G+ \\ &\qquad\qquad L_G^2\big\} \\ &= \sum_{k=0}^{\infty}\mathcal{H}\{N_0\}(k)\big\{k^2G^2+ \\ &\qquad\qquad k(2L_G+1)(G-1)G+ \\ &\qquad\qquad L_G(G-1)[L_G(G-1)+G]\big\}\end{aligned}$$

In terms of the values used in Equation (A.13), the expected value of k^2 over the pre-EM histogram is the variance of the pre-EM histogram $(V + \sigma_{\mathrm{CIC}})$ plus the square of the mean of the pre-EM histogram $(M + \sigma_{\mathrm{CIC}})$. Using these values, we obtain:

$$\begin{aligned}\left\langle j^2 \right\rangle = \; & \left(V + \sigma_{\mathrm{CIC}} + \left(M + \sigma_{\mathrm{CIC}}\right)^2\right) G^2 + \\ & \left(M + \sigma_{\mathrm{CIC}}\right)\left(2L_G + 1\right)\left(G - 1\right) G + \\ & L_G \left(G - 1\right)\left[L_G \left(G - 1\right) + G\right] +\end{aligned}$$

Now we compute the variance of the post-EM histogram, using the formula we have already derived for $\langle j \rangle$:

$$\begin{aligned}\left\langle j^2 \right\rangle - \left\langle j \right\rangle^2 = \; & \left(V + \sigma_{\mathrm{CIC}} + \left(M + \sigma_{\mathrm{CIC}}\right)^2\right) G^2 \\ & + \left(M + \sigma_{\mathrm{CIC}}\right)\left(2L_G + 1\right)\left(G^2 - G\right) \\ & L_G \left(G - 1\right)\left[L_G \left(G - 1\right) + G\right] \\ & - \left(M + \sigma_{\mathrm{CIC}}\right)^2 G^2 - 2\left(M + \sigma_{\mathrm{CIC}}\right)\left(G - 1\right) G L_G - \left(G - 1\right)^2 L_G^2 \\ = \; & \left(V + \sigma_{\mathrm{CIC}}\right) G^2 \\ & + \left(M + \sigma_{\mathrm{CIC}}\right)\left(G^2 - G\right) \\ & + L_G \left(G - 1\right) G\end{aligned}$$

Once we add the readout noise variance σ_{r} introduced by the readout stage, we have re-derived Equation (A.13).

Finally, we would like to point out some approximations and special cases of Equation (A.41). All of these are for $G \gg 1$. Since EM gains are typically $G \sim 1000$, these will be good approximations. The simplest approximation is to replace $(1 - 1/G)^{j-k}$ with $\exp(-\left(j - k\right)/G)$. The approximate error associated with this can be seen as follows, using the Taylor expansion of $\ln(1 + \epsilon) = \epsilon$ with an absolute error of at most $E(\epsilon) = \epsilon^2/2$:

$$\begin{aligned}
\left(1 - \frac{1}{G}\right)^{j-k} &= \exp\left((j-k)\ln\left(1 - \frac{1}{G}\right)\right) \\
&\approx \exp\left((j-k)\left(-\frac{1}{G} + \frac{1}{2G^2}\right)\right) \\
&= \left[\exp\left(-\frac{j-k}{G}\right)\right]\left[\exp\left(\frac{j-k}{2G^2}\right)\right] \\
&\approx \left[\exp\left(-\frac{j-k}{G}\right)\right] \qquad \text{(A.46)}
\end{aligned}$$

With $G \sim 1000$, the correction factor to $\exp(-(j-k)/G)$ is on the order of a percent for $j \sim 10^4$, which is the largest we are likely to see for our purposes.

The second approximation we will use is to assume that L_G and k are both of order unity or less. This is typically the case for dark frames with no signal. In this case, the histogram entering the EM stage is a Poissonian distribution with mean and variance σ_{CIC}, which is typically much smaller than 1 (unless we have a lot of thermal noise added to it). Thus, $\mathcal{H}\{N_0\}(k)$ is small enough to ignore unless k is very small. L is also typically small enough, and $L_G = L/\ln(G)$ is just a little smaller. For large gain, the histogram after the EM stage will be greatly spread out, with mean and variance on the order of G, so typically j values will be much larger than either k or L_G. These conditions allow us to make an approximation based on this identity:

$$\frac{\Gamma(j + L_G)}{\Gamma(j-k+1)} = \prod_{a=0}^{k-1}(j - k + L_G + a)\,\frac{\Gamma(j-k+L_G)}{\Gamma(j-k+1)}$$

where we used $\Gamma(x) = (x-1)\,\Gamma(x-1)$ repeatedly for the numerator. We now make the approximation that $j - k + L_G - a \approx j - k$, which is a decent approximation if j is larger compared to k and L_G:

$$\approx (j-k)^k\,\frac{\Gamma(j-k+L_G)}{\Gamma(j-k+1)} \qquad \text{(A.47)}$$

The percent error of the initial approximation, $j - k + L_G - a \approx j - k$ is on the order of $(L_G + k)/j$, which, under the conditions given above, is less than a percent.

Raised to the k^{th} power, for a small error, multiplies that error by k, which might bring the relative error up to a percent or so, which we deem acceptable.

Using the approximations in Equation (A.46) and Equation (A.47), we can write down an approximate form of the histogram after the EM stage given in Equation (A.41):

$$\mathcal{H}\{N_n\}(j) \approx \sum_{k=0}^{j} \frac{(j-k)^k\,\Gamma(j-k+L_G)}{\Gamma(k+L_G)\Gamma(j-k+1)} e^{-L} G^{-k} e^{-\frac{j-k}{G}} \mathcal{H}\{N_0\}(k) \tag{A.48}$$

for large gain G, small inputs k (and small leakage L and $L_G = L/\ln(G)$), and large outputs j.

The no-leakage case provides us with some insights for these histograms. If the leakage is $L = L_G = 0$, then the histogram after the EM stage given in Equation (A.41) can be written using combinatorial factors:

$$\mathcal{H}\{N_n\}(j) = \sum_{k=0}^{j} \binom{j-1}{k-1} G^{-k} \left(1-\frac{1}{G}\right)^{j-k} \mathcal{H}\{N_0\}(k) \tag{A.49}$$

If we apply the approximations in Equation (A.46) and Equation (A.47), this becomes:

$$\begin{aligned}\mathcal{H}\{N_n\}(j) &\approx \sum_{k=0}^{j} \frac{(j-k)^k\,(j-k-1)!}{(k-1)!\,(j-k)!} G^{-k} e^{-\frac{j-k}{G}} \mathcal{H}\{N_0\}(k) \\ &= \sum_{k=0}^{j} \frac{(j-k)^{k-1}}{(k-1)!} G^{-k} e^{-\frac{j-k}{G}} \mathcal{H}\{N_0\}(k) \end{aligned} \tag{A.50}$$

Note that, in this approximation, the histogram after the EM stage for a known number of electrons is just a shifted Poisson distribution, which gives some intuition on what the distributions should look like. We note in passing that the distribution for exactly one electron as input happens to look exponential, which is the first Poisson distribution. Once that is determined, the histogram for two input electrons is the convolution of the one-electron histogram with itself. Since each successive Poisson distribution can be generated by convolving one with the first Poisson distribution, the first Poisson distribution determines all the rest. Therefore, once we establish that the one-electron histogram is a lot like the first Poisson distribution, then the fact that the multiple-electron distributions look a lot like the higher Poisson distributions should be no surprise.

If we make a stronger small-noise approximation, we can produce a form that will give us a little more insight in the shape of histograms when there is no (or very little) leakage. If the CIC (and thermal) noise is sufficiently small, the histogram before the EM stage is well approximated by:

$$\mathcal{H}\{N_0\}(k) = \begin{cases} 1 - \sigma_{\text{CIC}} & \text{if } k = 0 \\ \sigma_{\text{CIC}} & \text{if } k = 1 \\ 0 & \text{if } k > 1 \end{cases}$$

This allows us to terminate the k in Equation (A.49) after $k = 1$. If $j = 0$, only the $k = 0$ term of the sum contributes, which results in just $\mathcal{H}\{N_0\}(0)$ ($\binom{-1}{-1} = 1$). This is because, when there is no leakage, you get zero electrons in the output only when (and every time that) no electrons enter the EM stage, so that histogram component is unchanged. For $j > 0$, only the $k = 1$ term of the sum contributes ($\binom{j-1\geq 0}{-1} = 0$), because you only get electrons in the output if there was an electron in the input to multiply. This allows us to write the output histogram as follows:

$$\mathcal{H}\{N_n\}(j) = \begin{cases} 1 - \sigma_{\text{CIC}} & \text{for } j = 0 \text{ (the } k = 0 \text{ term)} \\ \dfrac{\sigma_{\text{CIC}}}{G} e^{-\frac{j-1}{G}} & \text{for } j > 0 \text{ (the } k = 1 \text{ term)} \end{cases}$$

With these assumptions (large gain, no leakage, very little noise), we see that the output histogram of the EM stage is well-approximated by a delta-function peak at zero electrons with an exponential tail. What is more, using error functions, we can actually convolve this with the Gaussian from the readout noise. If the gain is large, we can assume these are continuous distributions in j, and that the single value at $j = 0$ is just a very narrow peak at $x = 0$, and the decaying exponential $\exp(-\left(j-1\right)/G)$ is simply $e^{-x/G}$ (the shift from $x - 1$ to x is assumed to be small enough to be not noticeable). We can then write the continuous histogram as (using $\delta(x)$ to represent a very sharp peak at $x = 0$ with area 1):

$$\mathcal{H}\{N_n\}(x) = \left(1 - \sigma_{\text{CIC}}\right)\delta(x) + \frac{\sigma_{\text{CIC}}}{G} e^{-\frac{x}{G}} \text{ for } x \geq 0 \qquad \text{(A.51)}$$

The convolution of a normalized Gaussian with the sharp peak is:

$$\frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} * \delta(x) = \int_{-\infty}^{\infty} \frac{e^{-\frac{(x-t)^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} \delta(t)\,\mathrm{d}t$$

$\delta(t)$ is assumed to be such a sharp peak that the Gaussian does not substantially change value during the time that $\delta(t)$ is non-zero, so we just replace the Gaussian with its $t = 0$ value:

$$\begin{aligned} &= \int_{-\infty}^{\infty} \frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} \delta(t)\,\mathrm{d}t \\ &= \frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} \int_{-\infty}^{\infty} \delta(t)\,\mathrm{d}t \end{aligned}$$

The integral of $\delta(t)$ is normalized to 1:

$$= \frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}}$$

Not surprisingly, this sharp peak just turns into a Gaussian when convolved with a Gaussian:

$$\frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} * \delta(x) = \frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} \tag{A.52}$$

The convolution of a Gaussian with an exponential is more difficult, but can be done if we use the error function, which is simply the integral of a Gaussian. We use this form:

$$\mathrm{erf}(x) := \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2}\,\mathrm{d}t$$

With this form, the error function is 0 at $x = 0$, and ± 1 at $x = \pm\infty$, and it is an odd function about $x = 0$. The exponential part is defined to be 0 for $x < 0$, so we restrict the integral in the convolution to $t > 0$:

$$\frac{e^{-\frac{x^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} * \frac{1}{G} e^{-\frac{x}{G}} = \int_0^{\infty} \frac{e^{-\frac{(x-t)^2}{2\sigma_\mathrm{r}^2}}}{\sqrt{2\pi\sigma_\mathrm{r}^2}} \frac{1}{G} e^{-\frac{t}{G}}\,\mathrm{d}t$$

Perform the usual trick of combining the exponentials and completing the square:

$$
\begin{aligned}
&= \frac{1}{G\sqrt{2\pi\sigma_{\mathrm{r}}^2}} \int_0^\infty e^{-\frac{1}{2\sigma_{\mathrm{r}}^2}\left[(x-t)^2+\frac{2\sigma_{\mathrm{r}}^2}{G}t\right]}\,\mathrm{d}t \\
&= \frac{1}{G\sqrt{2\pi\sigma_{\mathrm{r}}^2}} \int_0^\infty e^{-\frac{1}{2\sigma_{\mathrm{r}}^2}\left[t^2+2\left(\frac{\sigma_{\mathrm{r}}^2}{G}-x\right)t+x^2\right]}\,\mathrm{d}t \\
&= \frac{1}{G\sqrt{2\pi\sigma_{\mathrm{r}}^2}} \int_0^\infty e^{-\frac{1}{2\sigma_{\mathrm{r}}^2}\left[\left(t+\left(\frac{\sigma_{\mathrm{r}}^2}{G}-x\right)\right)^2+x^2-\left(\frac{\sigma_{\mathrm{r}}^2}{G}-x\right)^2\right]}\,\mathrm{d}t \\
&= \frac{1}{G\sqrt{2\pi\sigma_{\mathrm{r}}^2}} e^{\frac{\sigma_{\mathrm{r}}^2}{2G^2}} e^{-\frac{x}{G}} \int_0^\infty e^{-\frac{1}{2\sigma_{\mathrm{r}}^2}\left(t+\left(\frac{\sigma_{\mathrm{r}}^2}{G}-x\right)\right)^2}\,\mathrm{d}t
\end{aligned}
$$

Replace t with $t\sqrt{2\sigma_{\mathrm{r}}^2}$:

$$
\begin{aligned}
&= \frac{1}{G\sqrt{\pi}} e^{\frac{\sigma_{\mathrm{r}}^2}{2G^2}} e^{-\frac{x}{G}} \int_0^\infty e^{-\left(t+\left(\frac{\sigma_{\mathrm{r}}}{G\sqrt{2}}-\frac{x}{\sqrt{2\sigma_{\mathrm{r}}^2}}\right)\right)^2}\,\mathrm{d}t \\
&= \frac{1}{G} e^{\frac{\sigma_{\mathrm{r}}^2}{2G^2}} e^{-\frac{x}{G}} \frac{1}{2}\left\{\mathrm{erf}(\infty)-\mathrm{erf}\left(\frac{\sigma_{\mathrm{r}}}{G\sqrt{2}}-\frac{x}{\sqrt{2\sigma_{\mathrm{r}}^2}}\right)\right\}
\end{aligned}
$$

If we use $\mathrm{erf}(\infty)=1$, we get:

$$
\frac{e^{-\frac{x^2}{2\sigma_{\mathrm{r}}^2}}}{\sqrt{2\pi\sigma_{\mathrm{r}}^2}} * \frac{1}{G}e^{-\frac{x}{G}} = \frac{1}{G} e^{\frac{\sigma_{\mathrm{r}}^2}{2G^2}} e^{-\frac{x}{G}} \frac{1}{2}\left\{1-\mathrm{erf}\left(\frac{\sigma_{\mathrm{r}}}{G\sqrt{2}}-\frac{x}{\sqrt{2\sigma_{\mathrm{r}}^2}}\right)\right\}. \tag{A.53}
$$

This looks quite complicated, but it has a simple interpretation. There are some normalizing factors out front,[6] and the actual x dependence is $e^{-x/G}$ times half the quantity in braces. The quantity in braces produces a smooth ramp up from 0 to 2 as x changes from $-\infty$ ti ∞, with most of the change occurring within a few standard deviations (σ_{r}) from $x = 0$. This is from the Gaussian convolved with the jump from 0 to 1 that the exponential takes at $x = 0$ (because we define the actual function to be 0 for $x < 0$). Once you get past that ramp, the quantity in braces flattens out at 2 (which is multiplied by 1/2 to become 1), and the only remaining x dependence is an exponential decay with decay constant G, which is

[6] This function should be normalized to unity because it was the convolution of a normalized Gaussian with a normalized exponential, but you can actually integrate it and check by using integration by parts to reduce the error function to a Gaussian.

the same functional dependence as the original exponential. Thus, convolving an exponential with a Gaussian reproduces the same exponential, but with the initial jump smoothed out. In the limit of the gain G being much larger than the readout noise width σ_{r}, the prefactor becomes 1, and the original exponential is basically reproduced exactly, except for the smooth ramp.

Using the above results, we can actually perform the readout noise convolution on the histogram after the EM stage given in Equation (A.51). We will not write it out here, but it is just the sum of the convolution with a delta function (Equation (A.52)) weighted by $1 - \sigma_{\mathrm{CIC}}$ and the convolution of the exponential (Equation (A.53)) weighted by σ_{CIC}. The histogram is thus a Gaussian with an exponential tail. On a semi-log plot, this is a parabolic portion with a line tail on one side.

Fitting Dark-Frame Histograms to Measure EMCCD Noise Values

We will now discuss using variants of Equation (A.41) to fit actual dark-frame histograms to measure the noise levels of an EMCCD. The difficulty arises from the readout stage, which multiplies the electron count by an unknown gain factor and adds an offset (which can be determined). A typical EMCCD camera returns the readout for each cell as an unsigned integer, which is a non-negative integer. These integer readout values are usually termed *Analog Data Units*, or ADUs. Since some of the signal is actually negative (0 electrons is technically the minimum, but readout noise creates a range of outputs about that, which necessarily result in negative values), a constant offset is added to the signal so that 0 electrons maps to some other ADU value (typically a few thousand in the cameras we looked at). The readout stages also typically do not have unity gain, typically resulting in a single ADU corresponding to several electrons [102].

The easiest way to measure noise values is to look at dark frames, where light is prevented from reaching the EMCCD. The pre-EM histogram is then just a Poissonian distribution from the clock-induced charge (and thermal noise). For a decent EMCCD, with a short exposure and good cooling, the CIC and thermal noise will both be much less than 1, and the leakage should be fairly small as well. Under these conditions, Equation (A.51) gives a decent approximation for the histogram after the EM stage. The discussion following Equation (A.51) describes what the final output histogram should look like: a Gaussian peak with an exponential tail on one side. The width of the Gaussian peak gives the readout noise (in ADU, if we are looking at an actual histogram) while the center gives the ADU offset corresponding to 0 electrons, and the decay rate of the exponential tail gives the gain of the EM stage (scaled to ADU as well), while the area under the exponential tail gives us the CIC noise. We can therefore pretty much read off the offset and readout noise (in ADUs, not actual electrons), and a careful curve-fit gives us the gain (in ADUs) and combined CIC and thermal noise (in electrons). If we know the ADU/electron conversion, we will know the actual readout noise and gain; however, the actual value is actually not too important, since we will mostly care about how well we can distinguish the tail from the Gaussian. The relevant quantity for that is the ratio of the gain G to the readout noise σ_r, which remains the same whether we take the ratio of the actual values or the ADU values (which we get from the histogram). This can be seen by the fact that, in the convolution of Equation (A.51) with readout noise, using Equation (A.52) and Equation (A.53), both the ADU value x and the gain G are scaled by the readout noise σ_r.

For determining dark-counts, the actual scaling of the readout stage is a poor quantity to fit, because the quantity is almost redundant. For example, if we take the very-low-noise, no leakage, and large gain approximation used in Equation (A.51), the resulting histogram is such that there is *no way* to tell the difference between a histogram with a gain of $G = 1000$ and readout noise $\sigma_\mathrm{r} = 50$ e$^-$, and a different histogram with gain $G = 500$ and readout noise $\sigma_\mathrm{r} = 25$ e$^-$, if the latter has twice the gain in the readout stage and the CIC noises are the same. If you take the graph of the latter, and scale it horizontally by a factor of 2, the same result can be achieved by doubling the readout noise, since all x values in the equation are scaled by the readout noise. If we also scale the gain G by a factor of two, then every other change from the doubling of σ_r is undone, since all other occurrences are of the form σ_r/G. Technically, the two are slightly difference, because the exponential actually starts at $x = 1$ e$^-$ as opposed to $x = 0$ e$^-$, but if the readout noise is much larger than 1 e$^-$ (typical values are around 50 e$^-$), this change is lost in the much larger 0 e$^-$ Gaussian, and the convolution of the exponential with the same Gaussian.

Thus, to a reasonable approximation, we can eliminate this horizontal scaling as a fit parameter. In the very-low-noise, low-leakage, and high gain approximation, all the signal entering the EM stage has many fewer electrons than the readout noise. The histograms are well-described by Equation (A.48), once convolved with the readout noise. This equation was specifically written to emphasize that the sub-histograms for each input number of electrons k start at k—simply note that each occurrence of j is actually an occurrence of $j - k$, which essentially shifts the histogram over by k (this is valid for larger j, where the j in the upper limit of the sum is already much larger than the largest k for which there is a strong contribution, and so the upper limit may as well be ∞). Since we are going to convolve this with a wide Gaussian (compared to the range of k which contribute), these shifts will be insignificant, and so we ignore them by replacing $j - k$ with a continuous variable x (we will show this a little more carefully shortly). Our histogram (before readout noise convolution) then becomes:

$$\mathcal{H}\{N_n\}(x) \approx \sum_{k=0}^{\infty} \frac{\Gamma(x + L_G)e^{-L}}{\Gamma(k + L_G)\Gamma(x + 1)} \left(\frac{x}{G}\right)^k e^{-\frac{x}{G}} \mathcal{H}\{N_0\}(k) \tag{A.54}$$

which is a good approximation for when x, G, and σ_r are large compared to L, L_G, and all the k values for which the $\mathcal{H}\{N_0\}(k)$ inputs are large enough to be significant. If we use $\Gamma(x + 1 + L_G) = (x + L_G)\,\Gamma(x + L_G) \approx x\Gamma(x + L_G)$, which is true for $x \gg L_G$ (this is an approximation we could have lumped in with Equation (A.47) when we used it to derive this form) and use $e^{-L} = G^{-L_G}$ so we only have one leakage measurement, we can change this to:

$$\mathcal{H}\{N_n\}(x) \approx \sum_{k=0}^{\infty} \frac{\Gamma(x + 1 + L_G)G^{-L_G-1}}{\Gamma(k + L_G)\Gamma(x + 1)} \left(\frac{x}{G}\right)^{k-1} e^{-\frac{x}{G}} \mathcal{H}\{N_0\}(k) \tag{A.55}$$

The $k = 0$ case is very close to a delta function in x (and is one when $L_G = 0$), which is best seen in Equation (A.54). Note that the ratio $\Gamma(x + 1 + L_G)/\Gamma(x + 1)$ is 1 when $L_G = 0$, and, for small L_G, not much different from 1. The ratio is dependent on x, but will not change by large amounts if x changes by a factor of 5 or so (1 to 5 e$^-$/ADU being a typical range of readout stage scalings). In particular, even for a fairly large $L = 5\%$, and a range of x and G values from a few hundred to a few thousand, it changes by only a percent or so when x is changed by a factor of 10 or so in either direction. Given that the ratio changes very little, we are left with a sum of terms that are functions of x/G (or, for $k = 0$, a delta function in x). We could, therefore, write the convolution with the readout noise approximately as follows:

$$\mathcal{H}\{\text{final}\}(x) \approx \int_{-\infty}^{\infty} \frac{1}{\sigma_\mathrm{r}} G\left(\frac{y}{\sigma_\mathrm{r}}\right) \left(C_0 \delta(x-y) + f\left(\frac{x-y}{G}\right)\right) \mathrm{d}y$$

Here, we have used $G(y/\sigma_\mathrm{r})/\sigma_\mathrm{r}$ to represent a normalized Gaussian with the readout noise as the width, $C_0\delta(x)$ to represent the $k = 0$ term of Equation (A.55) (it is a delta function if $L_G = 0$, and is close otherwise), and $f(x/G)$ to represent the other terms of Equation (A.55). Earlier, we shifted various parts of the $f(x/G)$ terms by k. We could shift our integration variable y by the same amount, and the result is we are essentially convolving with a Gaussian shifted by k. As long as $k \ll \sigma_\mathrm{r}$ for all the important terms, the ratio of $G(y/\sigma_\mathrm{r})$ to $G((y-k)/\sigma_\mathrm{r})$ is $\exp(-yk/\sigma_\mathrm{r}^2)$ (neglecting a small $\exp(-k^2/2\sigma_\mathrm{r}^2)$ factor). Even for a fairly small readout noise of $\sigma_\mathrm{r} = 20$ e$^-$ (which is small for EM modes of operation), with a low CIC where $k = 2$ is as large as we need, this is less than a 3% correction for $|y| < \sigma_\mathrm{r}$ (which accounts for approximately 68% of the area under the Gaussian). Near the origin, this correction will get lost next to the large Gaussian from the $k = 0$ term (which was not shifted). Far from the origin, since G is typically on the order of several hundred or more, we can leave the shift in the $f(x/G)$ term, and use a Taylor expansion in k/G to second order, and look at the result. The zero-order term is the no-shift result and the first-order term integrates to zero (because it is an odd function). Thus, the error introduced to the shift is, to lowest order, proportional to $(k/G)^2$, multiplied by the second-order derivative of $f(x)$, multiplied again by the few-percent $k \ll \sigma_\mathrm{r}$ correction just computed. This justifies shifting the terms by k, which results in an insignificant error for small CIC and largish readout noise and gain.

Equation (A.55) is important because it allows us to fit a histogram without knowing the scaling performed by the readout stage. The $k = 0$ term is pretty close to a delta function, which, when convolved with the readout noise, gives a Gaussian with width σ_{r}. If you then scale that by some unknown scaling S, the result is a Gaussian with readout noise $S\sigma_{\mathrm{r}}$. For the $k > 0$ terms, as argued above, the x-dependence in the Gamma functions is very insensitive to scaling (and shifting) x, so we can treat them as functions of x/G. The convolution of these terms can then be written as:

$$\int_{-\infty}^{\infty} \frac{1}{\sigma_{\mathrm{r}}} G\left(\frac{y}{\sigma_{\mathrm{r}}}\right) f\left(\frac{x-y}{G}\right) \mathrm{d}y$$

If we scale x by S (by replacing x with x/S and renormalizing), the convolution looks like:

$$\int_{-\infty}^{\infty} \frac{1}{S\sigma_{\mathrm{r}}} G\left(\frac{y}{\sigma_{\mathrm{r}}}\right) f\left(\frac{x-Sy}{SG}\right) \mathrm{d}y$$

Replacing y with y/S everywhere gives:

$$\int_{-\infty}^{\infty} \frac{1}{S^2\sigma_{\mathrm{r}}} G\left(\frac{y}{S\sigma_{\mathrm{r}}}\right) f\left(\frac{x-y}{SG}\right) \mathrm{d}y$$

Here, we see this is exactly the same equation, except the readout noise and the EM gain have been scaled up by S (and the normalization is altered to compensate). This means that we can fit Equation (A.55) to a histogram in ADU instead of electrons, and the results are valid as long as we remember that the readout noise and EM gain have both been scaled by the readout stage scaling. That means we can correctly read off the CIC value and the ratio of the EM gain and readout noise (which causes the scaling to cancel) from the fit, without knowing the readout stage scaling factor. The leakage term L is still marginally affected (since we compute it by multiplying L_G by $\ln(G)$), but only logarithmically, so the correction tends to be small for large gain.

Thus, the fitting parameters we will use to fit to a dark frame are:

1. σ_{CIC}, the combined CIC and thermal noise

2. σ_{r}, the readout noise (in ADU)

3. G, the EM stage gain (in ADU)

4. L_G, the EM leakage rate divided by the natural log of G
5. the offset applied by the readout stage (in ADU)

The gain, by itself, is a unit-less number. The gain in ADU is simply the gain on electrons multiplied by the electron-to-ADU conversion factor (the readout stage scaling). Alternatively, it is the mean output signal value (in ADU) of a single input electron. The final formula is Equation (A.56) convolved with the readout noise Gaussian:

$$\mathcal{H}\{\text{final}\}(x) \approx \left\{ \frac{e^{-\dfrac{x^2}{2\sigma_\text{r}^2}}}{\sqrt{2\pi\sigma_\text{r}^2}} \right\} * \left\{ \sum_{k=0}^{\infty} \frac{\Gamma(x+1+L_G)G^{-L_G-1}}{\Gamma(k+L_G)\Gamma(x+1)} \left(\frac{x}{G}\right)^{k-1} e^{-\frac{x}{G}} \mathcal{H}\{N_0\}(k) \right\} \tag{A.56}$$

Although many approximations were used in deriving Equation (A.55), the initial model is consistent with how EMCCDs work, and all the approximations used were shown, or at least argued, to imply errors on the order of a few percent. One addition to the model was the leakage in the EM stage, which is at least consistent with the general model of how EMCCDs operate. The final test of whether or not this model is worthwhile is to compare it with actual histogram (it was also successfully compared with computer simulations which used the same basic model, but avoided some of the approximations), which is done in Chapter V. There, we show that this model, even with all its approximations, performs very well at fitting real EMCCD histograms, which gives us strong confidence that the parameters from a fit to a real histogram actually coincide with real noise measurements. There, we will also discuss how the model does not fit nearly so well without the leakage term, and discuss several other model alterations that, while sensible, fail to explain the discrepancy as well as adding the EM leakage to the model, which we take as strong evidence that the EM leakage is a real effect.

APPENDIX B

BASIC STATISTICAL RESULTS

This appendix is simply a collection of standard results in statistics and probability theory. The results and versions of the derivations may be found in many places, including many college-level probability textbooks, calculus textbooks, and statistical mechanics textbooks [107–109]. We include them here as a brief overview, a useful reference, and an introduction to the notation we use in this thesis. Most of the results in this chapter are simply-derived from first principles, which we often show as the techniques are useful and similar to some more complicated derivations used in other parts of this thesis.

General Statistics Formulas

The contents of this section are fairly general statistical results and definitions, and are included here mostly to nail down our notation and to provide us with some equations to reference, in case there is not a well-known name.

We often speak of a random variable with a particular probability distribution (or just of the distribution itself). If we have a random variable n, we would label its probability distribution $\mathcal{P}(n = m)$, or $\mathcal{P}_n(m)$, which is the probability than the random variable n is equal to the value m. Our distributions are typically discrete, and so we use discrete sum notation, but most of the results generalize to the continuous case by replacing the sum with an integral and the probability $\mathcal{P}_n(m)$ with a probability density multiplied by a differential (using the same notation, $\mathcal{P}_n(m)\,\mathrm{d}m$).

The expectation value is the expected value of a certain function of the random variable:

$$\langle f(n) \rangle = \sum_m \mathcal{P}(n = m) f(n). \tag{B.1}$$

where the sum is over all the allowed values of the random variable. The mean and variance are simply the expectation value of the random variable and the expectation value of the squared difference of the random variable from its mean, respectively:

$$\begin{aligned}\text{mean}(n) &= \langle n \rangle && \text{(B.2)}\\ \text{var}(n) &= \left\langle (n - \text{mean}(n))^2 \right\rangle \\ &= \left\langle n^2 - 2n\langle n \rangle + \langle n \rangle^2 \right\rangle = \left\langle n^2 \right\rangle - 2\langle n \rangle^2 + \langle n \rangle^2 = \\ &= \left\langle n^2 \right\rangle - \langle n \rangle^2 . && \text{(B.3)}\end{aligned}$$

The first and last forms of the variance provide two useful ways to compute the variance.

Often, we need to deal with a sum of multiple random variables, sometimes with their own unique distributions. When we have multiple random variables, we require a joint probability distribution. For example, if there are two random variables, n and m, then we would write $\mathcal{P}(n = n' \bigcap m = m')$ as the probability that $n = n'$ *and* $m = m'$. In this case, the expectation value of a function of these two random variables requires summing over two sets of possibilities:

$$\langle f(n,m) \rangle = \sum_{n'} \sum_{m'} \mathcal{P}\Big(n = n' \bigcap m = m'\Big) f(n', m').$$

Here, the first sum is over all possible values of n, and the second is over all possible values of m. When the function is simply the sum of two random variables, $n + m$, this reduces to $\langle n \rangle + \langle m \rangle$. This lets us write out the mean and variance of the sum of two random variables:

$$\begin{aligned}S &:= n + m \\ \text{mean}(S) &= \langle n + m \rangle = \langle n \rangle + \langle m \rangle = \text{mean}(n) + \text{mean}(m) && \text{(B.4)}\\ \text{var}(S) &= \left\langle (n + m - \text{mean}(n) - \text{mean}(m))^2 \right\rangle \\ &= \left\langle (n - \langle n \rangle)^2 + 2(n - \langle n \rangle)(m - \langle m \rangle) + (m - \langle m \rangle)^2 \right\rangle \\ &= \text{var}(n) + 2\langle nm - n\langle m \rangle - m\langle n \rangle + \langle n \rangle \langle m \rangle \rangle + \text{var}(m) \\ &= \text{var}(n) + 2(\langle nm \rangle - \langle n \rangle \langle m \rangle) + \text{var}(m). && \text{(B.5)}\end{aligned}$$

The means simply add, while variances add with an extra cross-term called the *covariance*.

Two random variables n and m are said to be *independent* if the following is true:

$$\mathcal{P}\Big(n = n' \bigcap m = m'\Big) = \mathcal{P}(n = n')\mathcal{P}(m = m'). \tag{B.6}$$

Essentially, this means that each random variable has its own probability distribution that does not depend on the particular value of the other random variable, and so the joint probability distribution is just the product of these two distributions. A quick example of two independent random variables would be two (ideal) coin flips, done with different coins, and recorded by separate observers. In such a case, the first variable is heads or tails completely independently of the value of the second flip (and vice versa). If the coin is fair, both flips have the same distribution (with a fifty-fifty probability), but that is not a requirement. In this example, the probability of having two heads is one quarter (one half times one half), as is the probability of the first flip being heads and the second being tails. A trivial example of non-independent random variables would be if the two observers recording the coin flips were actually recording the result of a *single* flip. In that case the result of the second recording would always be identical to the first recording. Both recordings have a fifty-fifty probability of being heads or tails, but the probability of having two heads is one-half, and the probability of having heads and tails is zero, since it is the same coin. Here, the probability of pairs of outcomes is *not* the product of individual outcomes.

In the case of independent random variables, the expectation value of the product splits:

$$\begin{aligned}
\langle nm \rangle &= \sum_{n'} \sum_{m'} \mathcal{P}\Big(n = n' \bigcap m = m'\Big) nm \\
&= \sum_{n'} \sum_{m'} \mathcal{P}(n = n')\mathcal{P}(m = m') nm \\
&= \left(\sum_{n'} \mathcal{P}(n = n') n\right) \left(\sum_{m'} \mathcal{P}(m = m') m\right) \\
&= \langle n \rangle \langle m \rangle .
\end{aligned} \tag{B.7}$$

This causes the covariance in Equation (B.5) to disappear, leaving us with these results for independent random variables:

$$S := n + m$$

$$\mathrm{mean}(S) = \mathrm{mean}(n) + \mathrm{mean}(m) \tag{B.8}$$

$$\mathrm{var}(S) = \mathrm{var}(n) + \mathrm{var}(m). \tag{B.9}$$

This easily generalizes to sums of many variables. The mean of the sum of random variables (independent or not) is the sum of the means of the individual variables, while the variance of the sum of *independent* random variables is the sum of the variances of each variables individually.

We sometimes need to work with *conditional probabilities.* Using our previous coin flip examples, if we see that one observer recorded heads, we might want to know the probability that the second observer recorded heads as well. In the independent example with two coins, that probability is one half. In the non-independent example where both observations were of the same coin, the probability is one. If our two random variables are n and m, we ask for the probability distribution of m (the second coin recording) *given* a particular value for n (the first coin recording), and write this conditional probability as:

$$\mathcal{P}(m = m' \mid n = n') = \frac{\mathcal{P}(n = n' \bigcap m = m')}{\sum_{m'} \mathcal{P}(n = n' \bigcap m = m')}. \tag{B.10}$$

The formula given above is how to compute a conditional probability. We first take the subset of all possible outcomes where the first random variable has the given value n'. The probability of being within that subset is the denominator of the right-hand-side of the above equation (where we simply took one term of what would have been a sum over all possible values for n). Within that subset, we take the probability where *both* $n = n'$ and $m = m'$ (the numerator), and compute the ratio of that probability compared to the total probability of the $n = n'$ subset. If n and m are independent variables, the Equation (B.10) rather trivially reduces to the single probability distribution for m.

Sometimes we encounter cases where it is easy to compute the conditional probability for m given a particular value for n, but it is not obvious how to compute the conditional probability for n given a particular value of m. This tends to happen when there is an obvious relation where the value of m is somehow caused by the result of n. For example, if we assign tails a value of 0 and heads a value of 1, and then perform two fair and ideal coin flips, we might ask what is the probability of the second flip being greater than the first, given that the first flip was heads (or tails). The answer is rather obviously zero if the first coins was heads, and one half if the first coin was tails. However, if we ask what the probability that the first coin was heads given that the second flip was NOT greater, that is a little more challenging. Here, we can simply write out all four possible outcomes, and compute the probability from that (2/3), but for more complicated problems, we can apply a more formal method.

This more formal method is a relatively straightforward extension of how we defined conditional probability in Equation (B.10). We start out by writing two conditional probabilities, one where we want the probability distribution for n given m, and the other where we want a probability for m given n:

$$\begin{aligned}\mathcal{P}(n = n' \mid m = m') &= \frac{\mathcal{P}(n = n' \bigcap m = m')}{\sum_{n'} \mathcal{P}(n = n' \bigcap m = m')} \\ \mathcal{P}(m = m' \mid n = n') &= \frac{\mathcal{P}(n = n' \bigcap m = m')}{\sum_{m'} \mathcal{P}(n = n' \bigcap m = m')}.\end{aligned}$$

We note that the numerators in these two equations are the same, and we can use that to substitute one conditional probability into the definition of the other:

$$\mathcal{P}(n = n' \mid m = m') = \frac{\mathcal{P}(m = m' \mid n = n') \sum_{m'} \mathcal{P}(n = n' \bigcap m = m')}{\sum_{n'} \mathcal{P}(n = n' \bigcap m = m')}. \qquad \text{(B.11)}$$

This result is called *Bayes's Theorem*, and is useful because it gives us a way to flip the conditional probability. In our above example, it is rather easy to compute the probability for m (whether the second flip was greater than the first) given n (the first flip), but more difficult to compute the other way around. Now, though, we can compute the other direction. The probability that $n = 1$ (the first flip was heads) given that $m = 0$ (the second flip was NOT greater) is:

$$\mathcal{P}(n = 1 \mid m = 0) = \frac{\mathcal{P}(m = 0 \mid n = 1) \sum_{m'} \mathcal{P}(n = 1 \bigcap m = m')}{\sum_{n'} \mathcal{P}(n = n' \bigcap m = 0)}.$$

Here, we can get the same 2/3 answer as before, but it does not save us any extra work over writing out the full table. In more complicated cases, however, having a formal method for relating the two conditional probability distributions comes in handy.

Sometimes we have one random variable that directly affects the outcome of a second random variable, but for some reason, not the converse (one may happen before the other, for instance). In cases such as these, it may be more convenient to use a conditional probability distribution rather than a joint probability distribution. As an example, if you are dealt an ace from a standard deck of cards (with a probability of 1/13), the probability that the second card is also an ace (4/51 if the first card was not an ace, and $3/51 = 1/17$ if the first card was an ace) is sometimes a little easier to describe as a conditional probability than a joint probability. If we know the conditional probability distribution for a random variable m (whether the second card is an ace) given some other random variables n (whether the first card is an ace), we can compute a probability distribution for the second random variable by substituting the conditional probability for the joint probability as given by Equation (B.10):

$$\mathcal{P}(m = m') = \sum_{n'} \mathcal{P}\Big(n = n' \bigcap m = m'\Big) = \sum_{n'} \mathcal{P}(m = m' \mid n = n')\mathcal{P}(n = n'). \tag{B.12}$$

For our two-card example, the probability that the second card is an ace is the probability that the first card is an ace multiplied by the probability that the second card is an ace given that ($1/13 \times 1/17$), plus the probability that the first card is not an ace multiplied by the probability that the second card is an ace given that ($12/13 \times 4/51$). Summing these together gives 1/13, which is correct, since if we ignore the first card, we might as well not have dealt it at all, and the probability of the second card being an ace is 1/13.

The two-card example above is a little simplistic (and is easier to solve without using conditional probabilities), but there is one common case where Equation (B.12) is actually an easier way to derive a formula. For example, if we know the probability distributions for two random variables n and m, we can derive the probability distributions for their sum. The only insight required is to notice that the probability that $n+m = S'$ given that $n = n'$ is exactly the same as the probability that $m = S' - n'$ (given that $n = n'$, if the two distributions are not independent):

$$\begin{aligned} \mathcal{P}(n+m = S') &= \sum_{n'} \mathcal{P}(n+m = S' \mid n = n')\mathcal{P}(n = n') \\ &= \sum_{n'} \mathcal{P}(m = S' - n' \mid n = n')\mathcal{P}(n = n'). \end{aligned} \tag{B.13}$$

In the case where the two probability distributions are independent, the conditional probability distribution for m reduces to simply the probability that $m = S' - n'$, and we see that the probability distribution of the sum of two independent random variables is the convolution of the individual probability distributions. In the case where the two probability distributions are not independent, this still looks like a convolution, but technically is not, because $\mathcal{P}(m = S' - n' \mid n = n')$ depends on n' as well as the difference $S' - n'$.

When we have the sum of two independent random variables, Equation (B.13) states that the probability distribution is the convolution of the individual probability distributions. Furthermore, Equation (B.8) and Equation (B.8) state that the means and variances add. We can easily check that any time we take the convolution of two normalized probability distributions, we get a normalized probability distribution where the sum and variances add. We start with a convolution of two arbitrary probability distributions, $\mathcal{P}_1$ and $\mathcal{P}_2$. The convolution is:

$$\mathcal{P}_1(n) * \mathcal{P}_2(n) = \sum_m \mathcal{P}_1(n-m)\mathcal{P}_2(m).$$

Since the convolution is a sum of non-negative quantities (the terms are products of probabilities, which need to be non-negative), it is itself a non-negative quantity.[1] The normalization is easy to compute if we simply substitute, switch the order of the sums, and shift one sum. We can always rearrange the terms like this because the sums of probabilities are absolutely convergent (they are always non-negative, and the total sum must be one). This works as long as we are not restricted to a

[1] Also, a quick change of variables to $n - m$ shows that the convolution is symmetric under a switch of $\mathcal{P}_1$ and $\mathcal{P}_2$.

finite interval, which is usually the case as we can always just define probabilities to be zero for the points outside the original allowed values:

$$\begin{aligned}\sum_n \mathcal{P}_1(n) * \mathcal{P}_2(n) &= \sum_n \sum_m \mathcal{P}_1(n-m)\mathcal{P}_2(m) \\ &= \sum_m \sum_{(n-m)} \mathcal{P}_1(n-m)\mathcal{P}_2(m) \\ &= \left(\sum_m \mathcal{P}_2(m)\right)\left(\sum_n \mathcal{P}_1(n)\right) = 1. \qquad \text{(B.14)}\end{aligned}$$

Up until the last equality, this is a general proof that the normalization of a convolution (of normalizable functions) is the product of the normalizations of the original functions. In the last equality, we assume the two original probability distributions are normalized, and so the convolution is also a proper distribution function. In fact, using the exact same steps as above, we can compute the expectation value of any function $f(n)$ under this convolved probability distribution (the above is a special case with $f(n) = 1$):

$$\begin{aligned}\sum_n \mathcal{P}_1(n) * \mathcal{P}_2(n) f(n) &= \sum_n \sum_m \mathcal{P}_1(n-m)\mathcal{P}_2(m) f(n) \\ &= \sum_m \sum_{(n-m)} \mathcal{P}_1(n-m)\mathcal{P}_2(m) f((n-m)+m) \\ &= \sum_m \sum_n \mathcal{P}_1(n)\mathcal{P}_2(m) f(n+m). \qquad \text{(B.15)}\end{aligned}$$

As expected, the last line is the expectation value for the function of a sum of two independent random variables, n and m (independent because the joint probability was factored into the product of single probability distributions, as in Equation (B.6)). Therefore, we can either insert $f(n) = n$ and $f(n) = (n - \langle n \rangle)^2$ to compute the mean and variance, respectively, of this convolved distribution, or we can simply jump straight to Equation (B.8) and Equation (B.9) to realize that the means and variances of convolved distributions simply add.

The Law of Large Numbers

There are several versions of the Law of Large Numbers, which effectively state that with a large number of independent measurements, the average of the measurements converges to the expected measurement value. A simple extension is that the expected deviations tend to decrease by the square root of the number of measurements.

We will simply demonstrate that the expected value of a sum of independent trials converges to the mean value of the underlying probability distribution. For any repeatable event, where each repetition is independent from all others, we can apply the facts that the mean and variances add (shown in Equation (B.8) and Equation (B.9)) to demonstrate these. If we perform N repetitions, with x_i being the value we measure on the i'th repetition, then, by Equation (B.8) (repeated many times):

$$\left\langle \sum_{i=1}^{N} x_i \right\rangle = N \left\langle x \right\rangle , \tag{B.16}$$

where $\langle x \rangle$ is the expected value for a single event, either computed from the known probability distribution for the events, or, typically, computed by the average of many measurements. If the distribution of x is localized enough that the variance is well-defined and finite, then, by Equation (B.9) (repeated many times):

$$\mathrm{var}\left(\sum_{i=1}^{N} x_i \right) = N \mathrm{var}(x), \tag{B.17}$$

where $\mathrm{var}(x)$ is the variance for a single event.

These equations mean that if some process is essentially composed of the sum of many independent processes, then the expected width of the distribution (usually taken as the square root of the variance, or the standard deviation) is the width of a single event multiplied by the square root of the number of measurements. If there are N underlying processes, then the distribution for the measurement will have a mean value of N times the mean for an individual process, and a standard deviation of $\sqrt{N}$ times the standard deviation for an individual process. The mean is the repeatable part of the experiment; the standard deviation quantifies how much the measurement changes each time. The ratio of the signal to noise is therefore $N/\sqrt{N} = \sqrt{N}$. Put another way, if we explicitly divide by N to attempt to measure the mean value for a single event, the noise in our measurement is $1/\sqrt{N}$.

This is where the oft-quoted $\sqrt{N}$ signal-to-noise ratio tends to come from. Some examples of such processes are flipping coins (adding up the number of heads), the integrated intensity from a low-intensity light, or the CIC noise in EMCCDs (these latter two are briefly mentioned in Section V.9). Low-intensity light tends to have a Poissonian distribution (covered in Section B.4), which tend to approximate many processes involving a low-likelihood of an event. Measuring the total power emitted over a small amount of time is like measuring the sum of multiple Poissonian distributions over short time intervals (which, as is described in Section B.4, is a Poissonian distribution with a larger mean). With a different description, a single atom has a certain probability of emitting during that time interval (or, if the interval is long enough, a probability distribution for emitting some number of photons), and the measurement is the sum of those distributions. If the measurement is repeated many times, the noise in the measurement will be $\sqrt{N}$ if the mean value of the signal if N photons.

We mentioned that the standard deviation is a measure of the width of a distribution. It is the square root of the variance, which is the average deviation (squared, to remove the sign) from the mean value. Another way to look at it is a bound for the distribution, where some fraction of the distribution is guaranteed to lie within a certain width. This bound, known as Chebyshev's Inequality, is not nearly as strong as the equivalent limits for a Gaussian distribution (mentioned in Section B.3), but holds for *all* distributions with a finite and well-defined standard deviation. A simple way to show this limit is to start with the definition of the variance (for simplicity, we assume the values have been shifted so that the mean value is zero):

$$\mathrm{var}(x) = \sum_x x^2 \mathcal{P}(x).$$

We then restrict the sum to only those values that are at least n standard deviations from the mean. Since we are dropping non-negative values, the result at most the same as the variance:

$$\mathrm{var}(x) \geq \sum_{|x| \geq n \mathrm{std}(x)} x^2 \mathcal{P}(x).$$

For this region, $x^2 \geq n^2\text{var}(x)$, so we can make another approximation:

$$\text{var}(x) \geq \sum_{|x| \geq n\text{std}(x)} n^2 \text{var}(x)\mathcal{P}(x).$$

Simplifying yields:

$$\frac{1}{n^2} \geq \sum_{|x| \geq n\text{std}(x)} \mathcal{P}(x). \tag{B.18}$$

This is essentially Chebyshev's Inequality. It is a statement that the probability of a value that is at least n standard deviations from the mean (the right-hand side) is at most $1/n^2$. Restating this result in terms of the probability of begin *within* n standard deviations (which is 1 minus the probability of not begin within) of the mean gives another version:

$$\mathcal{P}(x - \text{mean}(x)) < n\text{std}(x) > 1 - \frac{1}{n^2}.$$

This means that we can always infer that there is at least 75% of a distribution within 2 standard deviations of the mean, and larger fractions for slightly larger widths. In this sense, the standard deviation is a bound on the width of a distribution.[2]

The Central Limit Theorem

In physics, we often deal with Gaussian distributions. They are useful because they are smooth and localized functions, and they tend to show up often. The reason they show up often is described by the Central Limit Theorem. Proofs of this theorem usually involve concepts we do not use in this thesis, and so we will refer to other sources [107].

Informally, the central limit theorem states that the distribution of the sum of independent random variables with identical distributions approaches a Gaussian

[2] To show how this limit is as strong as possible, take a three-value distribution, where the probability of getting 0 is $1 - \epsilon$, and the probability of getting ± 1 is $\epsilon/2$. The mean is clearly zero, and the standard deviation is $\sqrt{\epsilon}$. In the limit $\epsilon \to 1$, the fraction within one standard deviation (just the 0 part, with probability $1 - \epsilon$) approaches 0. With $\epsilon = 1/4$, the fraction at least 2 standard deviations away (the ± 1 results, with probability ϵ) is 1/4, which is exactly the bound given by Chebyshev's Inequality.

distribution, as long as the individual distributions have a well-defined and finite variance. This is subject Equation (B.8) and Equation (B.9), so the mean and variance of the resulting Gaussian approximation is given by the mean and variance of the individual distributions, scaled by the number of distributions added. Therefore, the sum S of N identically-distributed independent random variables with mean $\text{mean}(x)$ and variance $\text{var}(x)$ can be approximated as a random variable with a Gaussian distribution with mean $N\text{mean}(x)$ and variance $N\text{var}(x)$. This approximation improves as N increases, and is often a good approximation for small-ish N ($N \sim 10$) around the mean of the final distribution. Deviations are often most apparent in the tails of the distribution.

If the individual random variables are discrete, the final sum will be as well, and while the final distribution will appear to have a Gaussian envelope, it will not be truly Gaussian because of that discreteness. The mean and variance of the distribution will be N times the individual mean $\text{mean}(x)$ and variance $\text{var}(x)$, and the formal central limit theorem states that if we subtract off the mean, and scale by the standard deviation, the resulting random variable $\left(S - N\text{mean}(x)\right)/\text{std}(x)\sqrt{N}$ will approach a Gaussian distribution with a mean of zero and a variance of unity. In the case of discrete inputs, the scaling causes the discreteness to compress into a continuous distribution as $N \to \infty$.

The Gaussian approximation suggested by the central limit theorem obeys the law of large numbers, but because the shape is known, we can put a much stricter constraint on Chebyshev's Inequality, given by Equation (B.18). We can actually compute the integral of the distribution outside n standard deviations for a Gaussian. The resulting function is often called the *error function*, and is often provided as a built-in function in numerical computation software packages. The precise normalization and offset of the error function is not standardized, but the computed results should be the same. We have found it useful to remember a few of the bounds, provided in Table B.1. We can clearly see that a Gaussian distribution is mcuh more restrictive than the Chebyshev bound for arbitrary distributions.

n	Gaussian	Chebyshev
1	31.7%	$\leq$ 100%
2	4.55%	$\leq$ 25%
3	0.270%	$\leq$ 11.1%
4	0.00633%	$\leq$ 6.25%

Table B.1. A comparison of Chebyshev bounds and exact values for Gaussian distributions. We provide some values for the probability that a random variable is at least n standard deviations from the mean for a Gaussian distribution, and compare that with the Chebyshev bound for arbitrary distributions, given by Equation (B.18).

Poisson Distributions

A common distribution in this thesis is the Possion distribution. Any time random, independent events happen with fixed average rate, the distribution of the number of events that happen in a particular time interval is a Poisson distribution. A process the produces such events, such as the clock-induced charge discussed in Chapter V, is called a *Poisson process*.

We will use Γ as the mean rate at which the events occur. For a very small time interval $\mathrm{d}t$, the probability of an event happening within that time is $\Gamma\,\mathrm{d}t$. If we write the number of events that occurs in a time interval of length t as n, we can write out a differential equation for the probability that no event occurs in that interval, $\mathcal{P}(n = 0, t)$:

$$\mathcal{P}(n = 0, t + \mathrm{d}t) = \mathcal{P}(n = 0, t)\,(1 - \Gamma\,\mathrm{d}t)\,. \tag{B.19}$$

This states that he probability that no events take place in a time interval $t + \mathrm{d}t$ is the probability that no events take place in the first time interval t *and* no events take place in the subsequent interval of length $\mathrm{d}t$ (one minus the probability that an event does happen). This is a simple differential equation to solve, since it basically states that the derivative of the function is proportional to the function. The standard solution is simply an exponential decay, scaled so that the probability of zero events in a time interval of length zero is unity:

$$\mathcal{P}(n=0,t) = e^{\Gamma t}. \tag{B.20}$$

From this, we can build up the probability of having larger number of events in a given time interval.

To compute the probability of having exactly one event in a time interval of length t, we first compute the probability that the event occurs at $t = t'$ with a small margin of error $\mathrm{d}t'$. The probability is equal to the probability of having no events from 0 to t', multiplied by the probability that the event occurs in the $\mathrm{d}t'$ interval, multiplied again by the probability that no events occur during the remaining $t - t'$ in the interval. Using Equation (B.20), that works out to be $e^{-\Gamma t'}\,(\Gamma\,\mathrm{d}t')\,e^{-\Gamma(t-t')} = e^{-\Gamma t}\Gamma\,\mathrm{d}t'$. If we then integrate t' over the entire interval to sum up the probability that the even happens somewhere within the interval, we compute the probability of having exactly one event:

$$\mathcal{P}(n=1,t) = \int_0^t e^{-\Gamma t}\Gamma\,\mathrm{d}t' = e^{-\Gamma t}\Gamma t. \tag{B.21}$$

This is easily generalized to multiple events. We pick multiple times for when each event happens, and multiply the probabilities that the events happen at exactly those times, with no events in between. The product of probabilities for having no events, being exponential in the time-interval, will multiply out to just $e^{-\Gamma t}$. The remaining probabilities work out to be Γ^n multiplied by all the small time intervals. Integrating all the possible times, keeping in mind that they all happen in order, yields the probability for having n events happening within a time interval of length t:

$$\mathcal{P}(n,t) = \frac{(\Gamma t)^n}{n!}e^{-\Gamma t}. \tag{B.22}$$

In the case of clock-induced charge, there is no time variable, but we could pretend that the excitation events happen over the time of a single clock cycle, or the entire shifting stage. It is apparent that the actual rate is not important, except when compared to the time interval. In Equation (B.22), we see this as the only dependence on rate and time occur through the product Γt. Therefore, if we halved the rate, but doubled the time interval, we would get exactly the same statistics. This lets us eliminate the time by defining $\lambda := \Gamma t$. We will see shortly that λ is the expected number of events, so we can define the Poisson distribution solely in terms of the expected number of events:

$$\mathcal{P}(n) = \frac{\lambda^n}{n!} e^{-\lambda}. \tag{B.23}$$

If we do have a events happening at random times with a fixed average rate Γ, then we simply need to remember that the expected number of events in a time interval t is $\lambda = \Gamma t$ to recover Equation (B.22). We also note that we can bypass using differentials by deriving this as a limiting case of a binomial distribution, which we will show in Section B.5.

We first compute some moments of the Poisson distribution. The easiest is to check the normalization:

$$\begin{aligned} \langle 1 \rangle &= \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} = 1. \end{aligned} \tag{B.24}$$

The sum is simply the Taylor series of an exponential, which cancels the exponential in Equation (B.23). The mean is also easy:

$$\begin{aligned} \langle n \rangle &= \sum_{n=0}^{\infty} n \frac{\lambda^n}{n!} e^{-\lambda} \\ &= e^{-\lambda} \sum_{n=1}^{\infty} \frac{\lambda^n}{(n-1)!} = \qquad (B.25) \\ &= e^{-\lambda} \lambda \sum_{n=0}^{\infty} \frac{\lambda^n}{n!} = e^{-\lambda} e^{\lambda} \lambda = \lambda. \qquad (B.26) \end{aligned}$$

Here, the extra n made the $n = 0$ term disappear, and cancelled part of the $n!$ in the denominator. We then pulled a factor of λ out, shifted the sum with $n \to n+1$, and then the sum was simply the Taylor expansion of an exponential again. We can use the same trick on $\langle n\,(n-1) \rangle$, which cancels two terms of the sum, and two factors in the factorial. We would then pull out λ^2, and then the sum could be shifted and would cancel the exponential. This yields $\langle n^2 - n \rangle = \lambda^2$. Using Equation (B.26), we find $\langle n^2 \rangle = \lambda^2 + \lambda$, and we can compute the variance using $\langle n^2 \rangle - \langle n \rangle^2$ from Equation (B.3):

$$\mathrm{var}(n) = \lambda. \tag{B.27}$$

For a Poisson distribution, the mean and variance are equal.

Since a Poisson process involves events happening randomly, if we have two sets of random events, we could combine them and the result would be another set of independent, random events. The number of events we would expect to see would then be the sum of the two counts we would expect from the individual processes. By this reasoning, the sum of two random variables with Poisson distributions should be have a Poisson distribution, with a mean equal to the sum of the previous means. This is easy to show. We start with two independent random variables n_1 and n_2 with Poisson distributions, and mean values λ_1 and λ_2, respectively. By Equation (B.13), the probability distribution for the sum n is given by:

$$\begin{aligned}
\sum_{n_2=0}^{n} \mathcal{P}(n_1 = n - n_2)\mathcal{P}(n_2) &= \sum_{n_2=0}^{n} \frac{\lambda_1^{(n-n_2)}}{(n-n_2)!} e^{-\lambda_1} \frac{\lambda_2^{n_2}}{n_2!} e^{-\lambda_2} \\
&= \frac{1}{n!} \sum_{n_2=0}^{n} \frac{n!}{n_2\,(n-n_2)!} \lambda_1^{(n-n_2)} \lambda_2^{n_2} e^{-(\lambda_1+\lambda_2)} \\
&= \frac{(\lambda_1+\lambda_2)^n}{n!} e^{-(\lambda_1+\lambda_2)}.
\end{aligned}$$

In the last step, we used the Binomial Theorem, or alternatively simply noted that the sum was exactly the expansion of $(\lambda_1 + \lambda_2)^n$. The resulting distribution is a Poisson distribution with mean $\lambda_1 + \lambda_2$, which is exactly what we wished to prove.

Since the sum of two Poissian random variables is also a Poissonian random variable, we can easily check some of our previous results. The means trivially add, as seen by our proof that the sum is Poissonian. Since the variances equal the means, the variances add as well, which means the law of large numbers holds as well. By induction, the sum of many Poissonian processes must also be Poissonian, which might seem to violate the central limit theorem, but this is not the case. Poissonian distributions, for mean values much larger than unity, are very similar to Gaussian distributions. Therefore, if we add together many Poissonian processes, the mean value will eventually become large compared to one, and the resulting distribution will be approximately Gaussian, and the central limit theorem does hold (if the means, and therefore variances, of the distributions are zero, then all of the distributions are Gaussian with zero width). By plotting Poissonian distributions with various mean values, and comparing them to Gaussian distributions with the same mean and variance, we see that even with mean values as low as 5, the Poissonian distribution looks qualitatively similar to the Gaussian distribution (except for the discreteness).

Often, Poissonian processes are used to describe rare events, such as the clock-induced charge (CIC) discussed in Chapter V. For very rare events, there are only two cases worth considering: zero or one events. As long as two events are too unlikely to concern us, we can model these cases as Poisson processes, since the distribution will match closely enough, with the right mean and variance. In our model, we consider the charge produced by a single shift to be Poissonian, if only because even a single charge is unlikely, and two or more are too unlikely to affect the statistics. Since each shift adds an independent amount of charge, the sum of all the shifts results in a Poissonian CIC distribution. However, even after all of the shifts, the mean value is still small (on the order of a few percent), and so even though we have added approximately one thousand Poissonian processes together, the result, while Poissonian, is definitely not Gaussian. This is an example of a case where the central limit theorem, while technically valid (since $1000 < \infty$), does not produce a useful approximation. The fact that Poissonian processes add to produce another Poissonian process is also why we are able to lump CIC and thermal noise together (along with any weak light signal, which may be Poissonian) in our discussion of EMCCDs.

As a final note, we discuss the case of photon absorption. Faint light tends to have a Poissonian distribution, and so should produce a Poissonian charge distribution when it hits an EMCCD. This should hold even though not every photon produces a charge. Assuming photon-to-charge conversion is independent of the amount of charge, then this is equivalent to taking a Poissonian distribution but randomly cutting out some of the events. The result is still a series of random, independent events, but with a reduced average rate. These are the only requirements for a Poissonian distribution, and so the result should be Poissonian. The effective rate is simply the number of events we would expect to get multiplied by the ratio of those events that actually generate charge.

Binomial Distributions

A binomial distribution results whenever some event with exactly one of two outcomes occurs multiple times. In general, when N events happen, each with one of two outcomes resulting with probability p, the number of events with that outcome, n, has a binomial distribution. The canonical example is flipping a coin and counting the number of heads that comes up. Here, N is the number of times the coin is flipped, p is 1/2 for a fair coin (the probability of getting heads), and n is the number of heads we get. To aid in our descriptions, we will refer to the two possible outcomes as successes or failures, with n being the total number of successes, even though often there is no strong distinction what should be a success and what should be a failure. We can write out the binomial distribution using simple combinatorics. The probability of success is p, and the probability of failure is $(1-p)$. The probability of a particular sequence of successes and failures is just the product of p for each success and $(1-p)$ for each failure. If we have N events, then the probability of a particular sequence with n successes is $p^n\,(1-p)^{N-n}$. The number of such sequences is $\binom{N}{n}$. All outcomes are equally likely, and so the probability of n successes is:

$$\mathcal{P}(n) = \binom{N}{n} p^n\,(1-p)^{N-n} = \frac{N!}{n!\,(N-n)!}\,(1-p)^{N-n}\,. \qquad \text{(B.28)}$$

Since we can think of this as the sum of single events, with either successes or failures as the outcome, we can also derive this as the sum of single events. Starting with a single event, Equation (B.28) reduces to a probability of p for a success and $(1-p)$ for a failure, which is the definition of p. Using an induction argument on N, we find that Equation (B.28) is the resulting probability distribution, using a recurrence relation on the binomial coefficients. This relation, $\binom{N}{n} = \binom{N-1}{n-1} + \binom{N-1}{n}$, is easy to prove by writing all the factors as fractions of factorials, and combining with common denominators. This argument also shows that the sum of binomial distributions, *if the probabilities of success are equal*, is also a binomial distribution. If we try to combine two binomially-distributed variables with different success probabilities, the result does not have a binomial distribution. A trivial way to see that is to add two cases, one where success is guaranteed ($p = 1$), and one where failure is guaranteed ($p = 0$). The result of these two trials is exactly one success and one failure. However, there is no value of p with $N = 2$ that will guarantee $n = 1$. Unless $p = 0$ or $p = 1$, n can be either 0, 1, or 2, with various non-zero probabilities. If $p = 0$ or $p = 1$, n will never be 1.

We will now derive the moments for a binomial distribution. We start by showing it is normalized:

$$\langle 1 \rangle = \sum_{n=0}^{N} \binom{N}{n} p^n \left(1-p\right)^{N-n} = \left(p - \left(1-p\right)\right)^N = 1^N = 1. \tag{B.29}$$

This follows from the Binomial Theorem, or, alternatively, realizing that the sum is exactly the expansion of $\left(a+b\right)^N$, with $a = p$ and $b = 1 - p$. Computing the mean uses the same trick, once we absorb the n factor into the binomial coefficient. This is done using $n\binom{N}{n} = N\binom{N-1}{n-1}$, which is easy to show by expanding the coefficients into factorials and cancelling terms. We avoid the $n = 0$ case by realizing that term is zero:

$$\begin{aligned}\langle n \rangle &= \sum_{n=0}^{N} n \binom{N}{n} p^n \left(1-p\right)^{N-n} \\ &= \sum_{n=1}^{N} N \binom{N-1}{n-1} p^n \left(1-p\right)^{N-n} \\ &= pN \sum_{n=1}^{N} \binom{N-1}{n-1} p^{n-1} \left(1-p\right)^{N-n} = pN. \end{aligned} \tag{B.30}$$

In the last step, we shifted the sum by letting $n \to n + 1$, which then turned the sum into the normalization of a different binomial distribution, which we solved in Equation (B.29). The result is not surprising. If there are N events, each with a success probability of p, we would expect pN successes. Using a similar trick twice, we find $\langle n\,(n-1)\rangle = p^2 N\,(N-1)$. Combining this with Equation (B.30), we find $\langle n^2\rangle = (pN)^2 + pN - p^2$. We can use that to compute the variance using $\langle n^2\rangle - \langle n\rangle^2$ from Equation (B.3):

$$\mathrm{var}(n) = pN\,(1-p)\,. \tag{B.31}$$

We see that both the mean and variance are simply the number of events N multiplied by the mean and variance for a single event, which we would expect because both the means and variances simply add for independent events.

We have already shown that the sum of multiple binomial random variables is also a binomial random variable, assuming that each has the same probability of success. As with our similar discussion for Poisson variables in Section B.4, the central limit theorem holds, which means that for a large enough number of trials, the binomial distribution should approach a Gaussian distribution. Also as with the Poissonian distribution, this happens once the mean value is distant from the endpoints, relative to the standard deviation. In the Poissonian case, there is only one endpoint, in that the number of counts cannot be negative. Here, the number of counts cannot be negative, but also cannot exceed N. So, as N becomes large, if p is neither 0 nor 1, then eventually the mean value pN will be far (in units of the standard deviation $\sqrt{pN\,(1-p)}$) from 0 and N, at which point, the distribution can appear Gaussian. This large separation is guaranteed, because while the separation grows proportional to N, the standard deviation only grows as $\sqrt{N}$. Again, like the Poissonian case, if $p = 0$ or $p = 1$, then the distribution remains an infinitely sharp peak, which can be thought of as a Gaussian with zero deviation. In the Poissonian case, we pointed out how the distribution appeared qualitatively Gaussian for fairly small mean values. The same applies here because, as we will show, the binomial distribution approaches a Poissonian in these cases.

We can use a binomial distribution to derive a Poissonian distribution. A Poissonian distribution is the distribution we expect if there are some random, independent events that occur with some average rate, so that we expect a certain number of events. We can think of this as having a very large number of events happening at a high constant rate, but where we only see the successes. In order to approximate a Poissonian process, we need to take the limit where the number of events goes to infinity, so that the successes may happen at any time, while simultaneously decreasing the probability of success so that the average number of successes, pN, If we take the binomial distribution, and take the $N \to \infty$ limit, while keeping $pN = \lambda$ fixed, our distribution is modified as follows:

$$\begin{aligned}\mathcal{P}(n) &= \binom{N}{n}\left(\frac{\lambda}{N}\right)^n\left(1-\frac{\lambda}{N}\right)^{N-n}\\ &= \binom{N}{n}\left(\frac{\lambda}{N\left(1-\lambda/N\right)}\right)^n\left(1-\frac{\lambda}{N}\right)^{N}\\ &= \binom{N}{n}\left(\frac{\lambda}{N-\lambda}\right)^n\left(1-\frac{\lambda}{N}\right)^{N}\end{aligned}$$

We compute the limit of this in the large N limit by taking a logarithm. Because the logarithm is continuous and monotonic, by the definition of continuity, if $\lim \ln(x) = \ln(y)$, then $\lim x = y$. After taking a logarithm and expanding the binomial coefficient in factorials, we find:

$$\ln(\mathcal{P}(n)) = \ln\left(\frac{N!}{(N-n)!}\right) - \ln(n!) + n\ln\left(\frac{\lambda}{N-\lambda}\right) + N\ln\left(1-\frac{\lambda}{N}\right)$$

The argument of the logarithm in the first term is the product of the n integers from N to $(N-n+1)$, inclusive. We can expand that logarithm of a product as the sum of logarithms of the individual factors,

$$\begin{aligned}\ln\left(\frac{N!}{(N-n)!}\right) &= n\ln(N)\\ &\quad +\ln(1) + \ln\left(1-\frac{1}{N}\right) + \cdots \ln\left(1-\frac{n-1}{N}\right).\end{aligned}$$

In the $N \to \infty$ limit, every term except the first approaches zero, so we keep only that term. We now have:

$$\ln(\mathcal{P}(n)) = n\ln(N) - \ln(n!) + n\ln\left(\frac{\lambda}{N-\lambda}\right) + N\ln\left(1-\frac{\lambda}{N}\right).$$

In the last term, expanding the logarithm yields $-\lambda/N + \mathcal{O}(N^{-2})$, which, when multiplied by the N factor, becomes $-\lambda$ plus terms that go to zero in the $N \to \infty$

limit. With that substitution, if we combine he first and third terms, our result is:

$$\ln(\mathcal{P}(n)) = n\ln\left(\frac{\lambda}{1-\lambda/N}\right) - \ln(n!) - \lambda$$

Now we can take $N \to \infty$, and when we invert the logarithm, we find the following result:

$$\mathcal{P}(n) = \frac{\lambda^n}{n!}e^{-\lambda}.$$

This is an alternative method to derive the Poisson distribution given in Equation (B.23).

REFERENCES CITED

[1] D. Budker and M. Romalis, Nature Physics **3**, 227 (2007).

[2] D. Melconian, J. A. Behr, D. Ashery, O. Aviv, P. G. Bricault, M. Dombsky, S. Fostner, A. Gorelov, S. Gu, V. Hanemaayer, et al., Phys. Lett. B **649**, 370 (2007).

[3] T. Rosenband, D. B. Hume, P. O. Schmidt, C. W. Chou, A. Brusch, L. Lorini, W. H. Oskay, R. E. Drullinger, T. M. Fortier, J. E. Stalnaker, et al., Science **319**, 1808 (2008).

[4] W. D. Phillips and H. Metcalf, Phys. Rev. Lett. **48**, 596 (1982), URL `http://link.aps.org/doi/10.1103/PhysRevLett.48.596`.

[5] S. Chu, L. Hollberg, J. E. Bjorkholm, A. Cable, and A. Ashkin, Phys. Rev. Lett. **55**, 48 (1985), URL `http://link.aps.org/doi/10.1103/PhysRevLett.55.48`.

[6] S. Chu, J. E. Bjorkholm, A. Ashkin, and A. Cable, Phys. Rev. Lett. **57**, 314 (1986), URL `http://link.aps.org/doi/10.1103/PhysRevLett.57.314`.

[7] E. L. Raab, M. Prentiss, A. Cable, S. Chu, and D. E. Pritchard, Phys. Rev. Lett. **59**, 2631 (1987), URL `http://link.aps.org/doi/10.1103/PhysRevLett.59.2631`.

[8] M. H. Anderson, J. R. Ensher, M. R. Matthews, C. E. Wieman, and E. A. Cornell, Science **259**, 198 (1995).

[9] K. B. Davis, M. O. Mewes, M. R. Andrews, N. J. van Druten, D. S. Durfee, D. M. Kurn, and W. Ketterle, Phys. Rev. Lett. **75**, 3969 (1995), URL `http://link.aps.org/doi/10.1103/PhysRevLett.75.3969`.

[10] W. Ketterle, M. R. Andrews, K. B. Davis, D. S. Durfee, D. M. Kurn, M. O. Mewes, and N. J. van Druten, Physica Scripta **T66**, 31 (1996).

[11] A. Öttl, S. Ritter, M. Köhl, and T. Esslinger, Phys. Rev. Lett. **95**, 090404 (2005), URL `http://link.aps.org/doi/10.1103/PhysRevLett.95.090404`.

[12] I. Bloch, J. Dalibard, and W. Zwerger, Rev. Mod. Phys. **80**, 885 (2008), URL `http://link.aps.org/doi/10.1103/RevModPhys.80.885`.

[13] H. Weimer, M. Müller, I. Lesanovsky, P. Zoller, and H. P. Büchler, Nature Physics **6**, 382 (2010).

[14] M. Lu, S. H. Youn, and B. L. Lev, Phys. Rev. Lett. **104**, 063001 (2010), URL `http://link.aps.org/doi/10.1103/PhysRevLett.104.063001`.

[15] H. K. Pechkis, D. Wang, Y. Huang, E. E. Eyler, P. L. Gould, W. C. Stwalley, and C. P. Koch, Phys. Rev. A **76**, 022504 (2007), URL `http://link.aps.org/doi/10.1103/PhysRevA.76.022504`.

[16] M. G. Raizen, A. M. Dudarev, Q. Niu, and N. J. Fisch, Phys. Rev. Lett. **94**, 053003 (2005).

[17] G. N. Price, S. T. Bannerman, K. Viering, E. Narevicius, and M. G. Raizen, Phys. Rev. Lett. **100**, 093004 (2008).

[18] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Phys. Rev. Lett. **100**, 240407 (2008).

[19] E. Narevicius, S. T. Bannerman, and M. G. Raizen, New J. Physics **11** (2009).

[20] J. J. Thorn, E. A. Schoene, T. Li, and D. A. Steck, Phys. Rev. A **79**, 063402 (2009), URL `http://link.aps.org/abstract/PRA/v79/e063402`.

[21] E. A. Schoene, J. J. Thorn, and D. A. Steck, Phys. Rev. A **82**, 023419 (2010), URL `http://link.aps.org/doi/10.1103/PhysRevA.82.023419`.

[22] A. Ruschhaupt, J. G. Muga, and M. G. Raizen, J. Phys. B: At. Mol. Opt. Phys. **39**, 3833 (2006).

[23] A. Ruschhaupt and J. G. Muga, Phys. Rev. A **70**, 061604(R) (2004).

[24] A. Ruschhaupt and J. G. Muga, Phys. Rev. A **73**, 013608 (2006).

[25] A. Ruschhaupt and J. G. Muga, Phys. Rev. A **76**, 013619 (2007).

[26] A. Ruschhaupt and J. G. Muga, J. Phys. B: At. Mol. Opt. Phys. **41**, 205503 (2008).

[27] B. T. Seaman, M. Krämer, D. Z. Anderson, and M. J. Holland, Phys. Rev. A **75**, 023615 (2007), URL `http://link.aps.org/doi/10.1103/PhysRevA.75.023615`.

[28] T. Bhattacharya, S. Habib, and K. Jacobs, Phys. Rev. Lett. **85**, 4852 (2000), URL `http://link.aps.org/doi/10.1103/PhysRevLett.85.4852`.

[29] S. Habib, T. Bhattacharya, A. Doherty, B. Greenbaum, K. Hopkins, A. Jacobs, H. Mabuchi, K. Schwab, K. Shizume, D. Steck, and B. Sundaram, arXiv (2005), URL `http://arxiv.org/abs/quant-ph/0505046v1`.

[30] J. B. Mackrory, K. Jacobs, and D. A. Steck, New J. Physics **12**, 113023 (2010).

[31] K. Jacobs and D. A. Steck, New J. Physics **13**, 013016 (2011).

[32] W. Rakreungdet, J. H. Lee, K. F. Lee, B. E. Mischuck, E. Montano, and P. Jessen, Phys. Rev. A **79**, 022316 (2009), URL `http://link.aps.org/doi/10.1103/PhysRevA.79.022316`.

[33] C. M. Trail, P. S. Jessen, and I. H. Deutsch, Phys. Rev. Lett. **105**, 193602 (2010), URL `http://link.aps.org/doi/10.1103/PhysRevLett.105.193602`.

[34] P. E. Gaskell, J. J. Thorn, S. Alba, and D. A. Steck, Rev. Sci. Inst. **80**, 1 (2009).

[35] P. Gaskell, URL `http://atomoptics.uoregon.edu/~zoinks`.

[36] T. Li, Ph.D. thesis, University of Oregon Department of Physics (2008).

[37] E. Schoene, Ph.D. thesis, University of Oregon Department of Physics (2010).

[38] *Hamamatsu em-ccd technical note* (2009), URL `http://sales.hamamatsu.com/assets/pdf/hpspdf/e_imagemtec.pdf`.

[39] *Princeton instruments proem brochure*, URL `http://www.princetoninstruments.com/pdfs/datasheets/princeton_instruments_ProEM_EMCCD_Brochure_RevA0.pdf`.

[40] R. Guntupalli, *Private communication*.

[41] *Low-light technical note 4, dark signal and clock-induced charge in l3vision*$^{\text{tm}}$ *ccd sensors* (2004).

[42] J. P. Gordon and A. Ashkin, Phys. Rev. A **21**, 1606 (1980), URL `http://link.aps.org/doi/10.1103/PhysRevA.21.1606`.

[43] P. D. Lett, R. N. Watts, C. I. Westbrook, W. D. Phillips, P. L. Gould, and H. J. Metcalf, Phys. Rev. Lett. **61**, 169 (1988), URL `http://link.aps.org/doi/10.1103/PhysRevLett.61.169`.

[44] D. J. Wineland and W. M. Itano, Phys. Rev. A **20**, 1521 (1979), URL `http://link.aps.org/doi/10.1103/PhysRevA.20.1521`.

[45] W. A. Hollerman (2002).

[46] F. Francisco, O. Bertolami, P. J. S. Gil, and J. Páramos, arXiv (2012), URL `http://arxiv.org/abs/1103.5222v2`.

[47] G. Barton, Journal of Physics B **7**, 2134 (1974).

[48] R. Loudon, *The Quantum Theory of Light* (Oxford University Press, 1983), 2nd ed.

[49] H. J. Metcalf and P. van der Straten, *Laser Cooling and Trapping* (Springer, 1999).

[50] C. E. Wieman and G. Flowers, Am. J. Phys. **63**, 317 (1995).

[51] D. A. Steck (2010), URL `http://steck.us/alkalidata`.

[52] K. I. Lee, J. A. Kim, H. R. Noh, and W. Jhe, Optics Letters **21**, 1177 (1996).

[53] J. A. Kim, K. I. Lee, H. R. Noh, and W. Jhe, Optics Letters **22**, 117 (1997).

[54] J. M. Kohel, J. Ramirez-Serrano, R. J. Thompson, L. Maleki, J. L. Bliss, and K. G. Libbrecht, J. Opt. Soc. Am. B **20**, 1161 (2003).

[55] R. S. Williamson III, P. A. Voytas, R. T. Newell, and T. Walker, Optics Express **3**, 111 (1998).

[56] J. Dalibard and C. Cohen-Tannoudji, J. Opt. Soc. Am. B **6**, 2023 (1989).

[57] A. Ashkin, Phys. Rev. Lett. **40**, 729 (1978), URL `http://link.aps.org/doi/10.1103/PhysRevLett.40.729`.

[58] A. Kastler, J. Opt. Soc. Am. **53**, 902 (1963).

[59] W. Ketterle, D. S. Durfee, and D. M. Stamper-Kurn, arXiv (1999), URL `http://arxiv.org/abs/cond-mat/9904034v2`.

[60] *Kodak kaf-0402 image sensor data sheet* (2010), URL `http://www.truesenseimaging.com/all/download/file?fid=8.49`.

[61] S. J. M. Kuppens, K. L. Corwin, K. W. Miller, T. E. Chupp, and C. E. Wieman, Phys. Rev. A **62**, 013406 (2000).

[62] J. F. O'Hanlon, *A User's Guide to Vacuum Technology* (John Wiley & Sons, Inc., 2003), 3rd ed., ISBN 0-471-27052-0.

[63] Y. T. Sasaki, J. Vac. Sci. Technol. A **9**, 2025 (1991).

[64] M. Bernardini, S. Braccini, R. De Salvo, A. Di Virgilio, A. Gaddi, A. Gennai, G. Genuini, A. Giazotto, G. Losurdo, H. B. Pan, et al., J. Vac. Sci. Technol. A **16**, 188 (1998).

[65] C. E. Wieman and L. Hollberg, Rev. Sci. Inst. **62**, 1 (1991).

[66] K. B. MacAdam, A. Steinbach, and C. Wieman, Am. J. Phys. **60**, 1098 (1992).

[67] L. Ricci, M. Weidemüller, T. Esslinger, A. Hemmerich, C. Zimmermann, V. Vuletic, W. König, and T. W. Hänsch, Optics Comm. **117**, 541 (1995).

[68] E. C. Cook, P. J. Martin, T. L. Brown-Heft, J. C. Garman, and D. A. Steck, Rev. Sci. Inst. **83**, 1 (2012).

[69] N. Bloembergen, Rev. Mod. Phys. **54**, 685 (1982), URL `http://link.aps.org/doi/10.1103/RevModPhys.54.685`.

[70] A. L. Schawlow, Rev. Mod. Phys. **54**, 697 (1982), URL `http://link.aps.org/doi/10.1103/RevModPhys.54.697`.

[71] P. G. Pappas, M. M. Burns, D. D. Hinshelwood, M. S. Feld, and D. E. Murnick, Phys. Rev. A **21**, 1955 (1980), URL `http://link.aps.org/doi/10.1103/PhysRevA.21.1955`.

[72] D. W. Preston, Am. J. Phys. **64**, 1432 (1996).

[73] C. Wieman and T. W. Hänsch, Phys. Rev. Lett. **36**, 1170 (1976), URL `http://link.aps.org/doi/10.1103/PhysRevLett.36.1170`.

[74] G. C. Bjorklund and M. D. Levenson, Phys. Rev. A **24**, 166 (1981), URL `http://link.aps.org/doi/10.1103/PhysRevA.24.166`.

[75] J. L. Hall, L. Hollberg, T. Baer, and H. G. Robinson, Appl. Phys. Lett. **39**, 680 (1981).

[76] H. L. Stover and W. H. Steier, Appl. Phys. Lett. **8**, 92 (1966).

[77] S. Kobayashi and T. Kimura, IEEE Journale of Quantum Electronics **QE-17**, 681 (1981).

[78] G. R. Hadley, IEEE Journale of Quantum Electronics **QE-22**, 419 (1986).

[79] P. Spano, S. Piazzolla, and M. Tamburrini, IEEE Journale of Quantum Electronics **QE-22**, 427 (1986).

[80] *Epo-tek 353nd data sheet.*

[81] E. W. Streed, Ph.D. thesis, Massachusetts Institute of Technology (2006).

[82] T. Meyrath, unpublished (2003), URL `http://george.ph.utexas.edu/~meyrath/informal/shutter.pdf`.

[83] K. Singer, S. Jochim, M. Mudrich, A. Mosk, and M. Weidemüller, Rev. Sci. Inst. **73**, 4402 (2002).

[84] D. Mitchell and P. Level, unpublished (2008), URL `http://www.phas.ubc.ca/~qdg/publications/InternalReports/LM-APSC479.pdf`.

[85] T. Meyrath and F. Schreck, *A laboratory control system for cold atom experiments: Hardware and software* (2009), URL `http://iqoqi006.uibk.ac.at/users/c704250/`.

[86] F. Jülicher, A. Ajdari, and J. Prost, Rev. Mod. Phys. **69**, 1269 (1997).

[87] H. Linke, M. T. Downton, and M. J. Zuckermann, Chaos **15**, 026111 (2005).

[88] S. Bize, M. S. Sortais, Y. Santos, C. Mandache, A. Clairon, and C. Salomon, Europhysics Letters **45** (1999).

[89] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C* (Cambridge University Press, 2002), 2nd ed., ISBN 0-521-43108-5.

[90] D. G. Zill and M. R. Cullen, *Differential Equations with Boundary-Value Problems* (Brooks/Cole, 2001), 5th ed., ISBN 0-534-38002-6.

[91] R. L. Burden and J. D. Faires, *Numerical Analysis* (Brooks/Cole, 2001), 7th ed., ISBN 0-534-38216-9.

[92] D. A. Steck (2012), URL `http://steck.us/teaching`.

[93] A. Gallagher and D. E. Pritchard, Phys. Rev. Lett. **63**, 957 (1989).

[94] P. D. Lett, K. Mølmer, S. D. Gensemer, K. Y. N. Tan, A. Kumarakrishnan, C. D. Wallace, and P. L. Gould, J. Phys. B: At. Mol. Opt. Phys. **28**, 65 (1995).

[95] G. R. Fowles and G. L. Cassiday, *Analytical Mechanics* (Saunders College Publishing, 1990), 6th ed., ISBN 0-03-022317-2.

[96] J. V. José and E. J. Saletan, *Classical Dynamics: A Contemporary Approach* (Cambridge University Press, 2000), ISBN 0-521-63636-1.

[97] J. C. Maxwell, *Theory of Heat* (Longmans, Green, and Co., New York, 1871).

[98] C. H. Bennett, Scientific American **257**, 108 (1987).

[99] R. J. Scully and M. O. Scully, *The Demon and the Quantum* (Wiley-VCH, Weinheim, 2007).

[100] L. D. Landau and E. M. Lifshitz, *Statistical Physics* (Gopsons Papers Ltd., 2002), 3rd ed.

[101] *Hamamatsu imagem data sheet*.

[102] *Princeton instruments proem:512bk data sheet*, URL `http://www.princetoninstruments.com/Uploads/Princeton/Documents/Datasheets/Princeton_Instruments_ProEM_512BK_eXcelon_rev_M3.pdf`.

[103] R. W. Engstrom, *Photomultiplier Handbook* (RCA/Burle, 1980).

[104] C. W. Helstrom, J. Appl. Phys. **55**, 2786 (1984).

[105] *Andor ixon 897 data sheet*, URL `http://www.andor.com/pdfs/specifications/Andor_iXon_897_Specifications.pdf`.

[106] *Photometrics evolve:512 data sheet*, URL `http://www.photometrics.com/products/datasheets/evolve_512.pdf`.

[107] I. Miller and M. Miller, *John E. Freund's Mathematical Statistics* (Prentice Hall, 1999), 6th ed.

[108] J. Stewart, *Calculus* (Brooks/Cole, 1995), ISBN 0-534-25158-7.

[109] D. V. Schroeder, *An Introduction to Thermal Physics* (Addison Wesley Longman, 2000), ISBN 0-201-38027-7.