Optical/UV Single-Photon Spectrometers using Superconducting Tunnel Junctions

A Dissertation
Presented to the Faculty of the Graduate School
of
Yale University
in Candidacy for the Degree of
Doctor of Philosophy

by

Christopher Mogan Wilson, Jr.

December 2002

Yale University
New Haven, Connecticut

Abstract

Optical/UV Single-Photon Spectrometers using Superconducting Tunnel Junctions

Christopher Mogan Wilson, Jr.
Yale University
2002

We report progress on the ongoing development of optical/UV single-photon, spectrometers using superconducting tunnel junctions. Our devices utilize a lateral trapping geometry. Photons are absorbed in a Ta thin film, creating excess quasiparticles. Quasiparticles diffuse and are trapped by Al/AlOx/Al tunnel junctions located on the sides of the absorber. The Ta/Al interface does not overlap the junction area. Devices designed for imaging have tunnel junctions on two opposite sides of the absorber. Position information is obtained from the fraction of the total charge collected by each junction. Using devices designed for large backtunneling gain, we have measured an energy resolution of 0.4 eV at 4.89 eV. The resolution in these devices is limited by thermodynamic fluctuations of the thermal quasiparticle number in the Al trapping layers. We predict that this previously unconsidered noise source should be important in any backtunneling device with "deep" traps. We also report preliminary measurements of a second generation of detectors designed to eliminate this noise and, consequently, backtunneling gain. These devices need small junctions, of order 1 μm^2, for best noise performance. With currently available amplifiers, the resolution of these devices should approach the intrinsic limits of creation and trapping statistics, and exceed the resolution of devices with backtunneling.

Table of Contents

Chapter 1: Introduction

Chapter 2: Experimental Apparatus and Conditions

Chapter 3: Physical Noise Processes--Thermodynamic Fluctuations

Chapter 4: Physical Noise Processes--Master Equation Applications

Chapter 5: Device Performance

List of Figures

List of Tables

Symbols and Abbreviations

Vectors and matrices indicated by **bold** characters in the text.

α, β	constants
γ	average photon number
Γ	rate constant
Δ	superconducting energy gap
ε	average energy required to break a Cooper pair
ε_F	Fermi energy
$\Phi(u)$	correlation function
λ	photon wavelength
ω	angular frequency
σ^2	variance
τ	time constant
AC	alternating current (signal frequency)
A_j	tunnel junction area
Al	aluminum
$\mathbf{a}$	occupation vector
$\mathbf{B}$	second order Fokker-Planck moments
BNC	bayonet N-connector
C	capacitance
CCD	charge coupled device
$D(\varepsilon_F)$	two-spin density of states at the Fermi energy
DC	direct current (low frequency)
E	energy
E_ω	phonon energy
E_γ	photon energy
e-	electron's charge
ESA	European Space Administration
eV	electron volt
e_n^2	voltage-noise power spectral density
f	fraction
f	frequency
FET	field-effect transistor
FFT	fast Fourier transform
F	Farad
F	Fano factor
F′	effective Fano factor
F_ω	phonon trapping factor
$\mathbf{G}(\omega)$	cross-power spectra matrix
g(N)	generation parameter
^{3}He	isotope of helium with atomic mass of 3

^{4}He	isotope helium with atomic mass of 4
Hz	Hertz
I	current
i_n^2	current-noise power spectral density
i,j,k	counting variables
K	Kelvin
k_B	Boltzmann's constant
L	absorber length
ΔL	pixel size
M	linearized transition matrix
N	occupation number
N^0	steady-state occupation number
ΔN	variation of occupation number
δN	transition shot size
N_γ	number of photon-excited quasiparticles
Nb	niobium
ND	neutral density
O	oxygen
$\mathcal{O}(\ )$	order of magnitude
P( )	probability function
p_{ij}	two-level transition probability
PMT	photomultiplier tube
Q_γ	charge signal created by photon
q_{eff}	effective charge
$Q(\mathbf{a};\mathbf{a}')$	multilevel transition probability
R	resistance
R_{NN}	normal-state resistance of a tunnel junction
R	responsivity
R	recombination constant
R*	effective recombination constant
r(N)	recombination parameter
Si	silicon
STJ	superconducting tunnel junction detector
S(ω)	power spectral density
S	entropy
$\mathbf{s}^0$	matrix of entropy second derivatives
T	temperature
Ta	tantalum
TES	transition-edge sensor detector
UV	ultraviolet
V	voltage
Vol,V	volume
X_k	intrinsic thermodynamic variable
Z	impedance

Acknowledgements

The work presented in this dissertation has been a collaborative effort in many ways. First, I would like to thank my advisor, Daniel Prober. I could not have asked for a better advisor. He has been supportive of myself and of my efforts. He has encouraged my pursuits, when appropriate, and discouraged them, when appropriate. He has helped me develop an appreciation for the practical aspects of science, teaching me that applications are not a bad thing. He has also consistently shown that he has the best interests of his students at heart. I am very grateful to him.

I would also like to thank the other professors that have helped me along the way. Peter Kindlmann has provided invaluable advice on electronics. Robert Wheeler has always provided a fresh perspective and some history. Robert Grober and Robert Schoelkopf have often provided a skeptical though helpful point of view. Michel Devoret has offered a subtle physical insight in recent years and Andrew Szymkowiak has shared a wealth of experience. Finally, Douglas Stone has also been supportive and interested in my work.

I worked with a number of graduate students and post-docs, directly and indirectly. I would like to thank Michael Gaidis and Stephan Friedrich for all the great work they did before I started on the project. Stephan and I have become good friends through our brief interaction at Yale and our many meetings at conferences. I have worked most closely with Ken Segall and Luigi Frunzio. Ken and I worked together with an ease and rapport that I fear will be hard to find again. We were always good at listening to each other. I have spent many long hours in the clean room and the laboratory with Luigi. He has boundless energy and an amazing attention to detail. I was pleased to attend his wedding in Italy in June 2001 and glad that he could return to New Haven to attend mine in May 2002. I have worked along side Liqun Li for a number of years. She has provided consistent help and taught me to be patient. I have also been helped by a number of other students including Irfan Siddiqi, Minghao Shen, Ryan Held, Wenzjuan Zhu, Andrea Davies, Ben Mazin, and Dave Schiminovich. Finally, I have enjoyed working with and directing a number of undergraduates, including Rich Lathrop, Jeff Perlman, Aryesh Mukherjee and Avi Robinson-Mosher.

I would like to thank all my friends and family for being so supportive along the way. Both of my parents have doctorates in physical sciences, and although they have never pressured me, I think they are happy to see me complete my PhD. While I have attended Yale, I have also earned a black belt in Aikido. I would like to thank all of my friends at the dojo for providing prospective and sanity. I would also like to thank all of my friends in the Midsummer diaspora. They have kept me grounded all these years and added meaning to my endeavors. Most of all, I would like to thank Melissa, my wife of two months. We met our first year in that dreary graduate school dorm. She has been by my side the whole way and, without her, this would not have been possible.

Chapter 1: Introduction

1.1 Motivation

Optical and ultraviolet (UV) photon detectors based on semiconductor technologies have advanced significantly in recent years. Various types of detectors can provide spectroscopy, imaging, single-photon sensitivity, photon timing or high quantum efficiency. However, no single semiconductor detector can provide all of these things. While many applications only need a detector with one or two of the above characteristics, the next generation of applications in many fields will demand detectors that combine many of them. No semiconductor technology promises to combine all of these characteristics in a single detector. However, a new class of cryogenic detectors based on superconducting technology does promise such "all-in-one" detectors [Peacock 1996], [Cabrera 1998].

This dissertation presents research into one type of cryogenic detector, the superconducting tunnel junction (STJ) detector. The basic operating principle is similar to many semiconductor detectors: an incident photon deposits energy in the detector that excites carriers which then produce a measurable electrical signal. The major difference between a semiconductor and superconducting detector is the minimum amount of energy required to produce a single excitation, often called the energy gap. In most semiconductors, the energy gap is of order 1 eV. In conventional superconductors, the energy gap is of order 1 meV. This implies that the same amount of energy absorbed in a superconductor will produce approximately 1000 times the number of excitation as in a semiconductor! This fact leads to a number of practical advantages. In particular, it means that the energy resolution of STJ detectors can greatly exceed that of semiconductor detectors. Typically, the resolving power of a detector based on carrier excitation is proportional to the square root of the average number of excitations created. (The resolving power, R, is the energy of the absorbed photon divided by the uncertainty in the detected energy.) For example, photons in the optical range have an energy between about 1.8 eV and 3 eV. Thus, an optical photon will produce about 1 excitation in a semiconductor regardless of the energy of the photon. Therefore, the relative uncertainty in the energy is large, of order the energy of the photon itself. On the other hand, the same photon will produce thousands of excitations in an STJ, and we will be able to distinguish photons of different energies.

STJ detectors are not the only class of superconducting spectrometers. Microcalorimeters are a broad class of detectors based on detecting the increase in temperature of a small absorber caused by the absorption of a single photon. The increase in the absorber's temperature can be measured in a number of the ways, but in the optical/UV energy range, a transition edge sensor (TES) is used. A TES is a strip of superconducting metal held in its superconducting transition by electrothermal feedback. In the transition region, the resistance of the TES is a strong function of temperature, making the TES a very sensitive thermometer.

Intrinsic energy resolution across the optical/UV spectrum is the main motivation for developing superconducting detectors. A number of applications in astronomy and biology could benefit from a optical/UV detector that combines energy resolution with other characteristics. One example from astronomy would be the mapping of the universe in three dimensions (3D). When a semiconductor detector, such as a charge coupled device (CCD), is exposed to the sky it produces a 2D image. In general, the image contains no independent information about how far away any object in the image is. If an imaging array of STJ detectors were exposed to the same part of the sky, it would not only record the image, but also an energy spectrum from each object. This energy spectrum can be used to determine the red shift of each object [de Bruijne 2002], from which its distance from Earth can be inferred according to Hubble's Law. The astronomical image produced by the STJ is therefore a real 3D image of the sky. The same information could be achieved by exposing the CCD many times with different color filters, but this is not always practical. If the objects being recorded are very faint, many exposure may require an unworkable amount of observation time. For example, the famous Hubble Deep Field image [Williams 1996] was produced by exposing its CCD three times with different color filters (enough resolution to print the image in color). With each exposure requiring an entire observation, extracting more spectra would have been too expensive. An STJ camera, on the other hand, could produce the equivalent of 20-50 CCD exposures in a single observation.

Other astronomical applications include recording time-resolved spectra of variable sources. Compact sources that vary rapidly in time, such as pulsars and cataclysmic binaries, have recently begun to be studied with cryogenic detectors[Verhoeve 2002], [Cabrera 2002]. For the first time, cryogenic detectors have recorded time-resolved, broad-band spectra of these objects, providing new information on their structure and dynamics. The introduction of superconducting detectors with this ability is opening new areas of exploration for this broad class of astronomical sources.

There are also a host of potential applications of cryogenic detectors in the fields of single molecule physics and biology. In many of these applications, fluorescent molecules are introduced into a system as probes. These probes often have absorption and emission spectra that change in response to environmental properties such as pH or the presence of certain ions. Many experiments done now are essential static experiments or they look at changes only on second or minute time scales. However, many physical time scales in the systems are much faster. Processes such as diffusion of molecules or binding and unbinding can take place on time scales of order 1 μs or less. Any information related to these fast dynamics is lost to conventional detectors. STJ detectors promise to resolve spectra on these short time scales, providing a wealth of new information about the systems and the physics of the probes themselves.

1.2 Operating Principle

The basic element of an STJ detector is the superconducting tunnel junction, comprised of two superconducting electrodes separated by a very thin insulating barrier. In our detectors, the electrodes are made from aluminum (Al) and the insulating layer is native Al-oxide (Al_20_3). By applying a magnetic field in the plane of the barrier, we can suppress the critical current of the junction to nearly zero. (The critical current is the maximum Cooper pair current that can flow before a finite voltage develops across the barrier.) We can then bias the junction at a finite voltage. At zero temperature, all electrons in the system would be condensed as Cooper pairs and there would be no current in the subgap region, i.e., for bias voltages less then the gap voltage of 2Δ/e- (where Δ is the superconducting energy gap of the electrodes and e- is the electron's charge). When a single photon is absorbed in one of the electrodes it breaks Cooper pairs, creating excess quasiparticles. These quasiparticles can then tunnel across the barrier, producing a pulse of excess subgap current. We measure this pulse and then integrate it to obtain the total tunneled charge, which tells us how many excess quasiparticles were created. The average number of quasiparticles created, N_γ and therefore the charge collected at high bias voltage, Q_γ, is related to energy of the incident photon, E_γ, by the relation

$$N_\gamma = \frac{Q_\gamma}{e^-} = \frac{E_\gamma}{\varepsilon} \tag{1.1}$$

where ε is the average energy required to create a single quasiparticle. Monte Carlo simulations of the absorption process predict that ε = 1.7Δ [Rando 1992]. The constant ε is larger than Δ because some of the photon's energy escapes the electrodes as phonons with an energy too small to break Cooper pairs.

If we repeat the energy measurement on many photons of identical energy, we will find a distribution of collected charges. The full width at half maximum (FWHM) of this distribution determines the resolution of our energy measurement. Many things can contribute to the width of the distribution, but the width is fundamentally limited by the processes that create quasiparticles after the absorption of the incident photon. These processes, such as phonon emission by energetic quasiparticles, are all random. This leads to an intrinsic variation in the number of quasiparticles created. The minimum energy resolution dictated by this variation is

$$\Delta E_{Fano} = E_\gamma \frac{\Delta N_\gamma}{N_\gamma} = 2.355\sqrt{F\varepsilon E_\gamma} \tag{1.2}$$

where F is the Fano factor and the factor of 2.355 converts from one standard deviation of noise to FWHM. This intrinsic limit is often called the Fano limit. The Fano factor accounts for correlations in the creation processes and is calculated to be F = 0.2 [Kurakado 1982].

We fabricate the electrodes of our junctions from Al because Al-oxide naturally forms a very high quality tunnel barrier. However, Al is a poor photon absorber: in the optical/UV range it is highly reflective and in the X-ray range it has a small absorption cross section. Also, we typically want the volume of the STJ electrodes to be small so that quasiparticles tunnel quickly before they are lost. On the other hand, we want the detector itself to have a large volume, so that it collects more photons. Our solution to these engineering conflicts is to use an absorber that is distinct from the tunnel junction, but in good metallic contact with it. We make absorbers from tantalum (Ta) which has a large absorption cross section in both the optical/UV [Verhoeve 1997] and X-ray range [Henke 1993]. Using a Ta absorber also solves the volume "problem" through a process known as quasiparticle trapping. The energy gap of Ta is $\Delta_{Ta} = 700$ μeV, about four times larger than the energy gap of Al, $\Delta_{Al} = 180$ μeV. This means that the minimum energy (above the Fermi energy) at which a quasiparticle can exist in the Al is much less than the minimum energy in Ta. Thus, when excess quasiparticles are created in the Ta absorber, they can diffuse freely until they enter the Al electrode. When they enter the Al, they are at a relatively high energy and they typically scatter inelastically very quickly, emitting phonons and

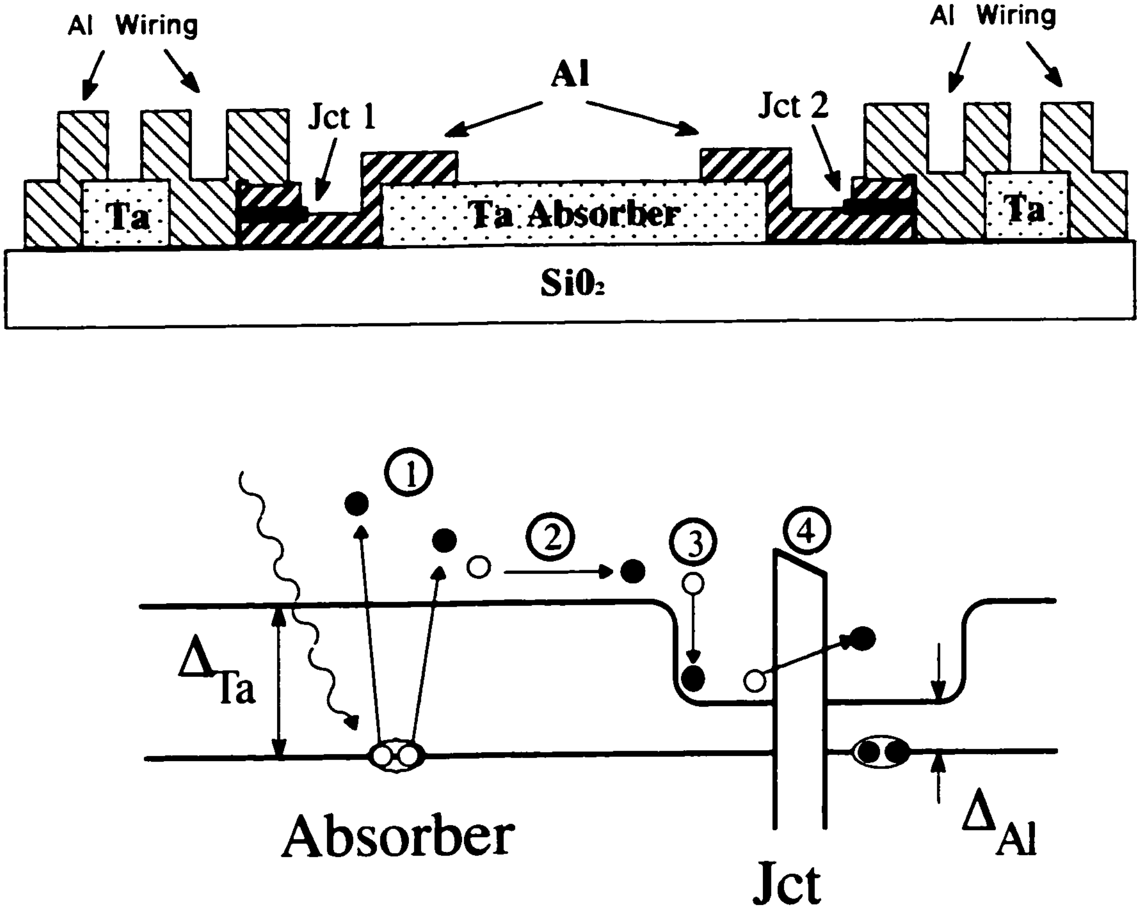


Figure 1.1 Schematic of an imaging STJ detector (upper) and a energy band diagram showing one side of the device (lower). Not shown is an insulating SiO layer between the trap and wiring, which has a via (hole) for contact between the two layers. The T a plugs interupting the Al wiring are omitted for devices designed without backtunneling. Physical processes illustrated in the band diagram are: 1) quasiparticle creation, 2) diffusion, 3) trapping and 4) tunneling.

losing energy until they approach Δ_{Al} (see Fig. 1.1). Once the quasiparticles have scattered to an energy less than Δ_{Ta}, they cannot diffuse back into the Ta absorber because there are no allowed states. Once the quasiparticles are "trappedÓ in the Al electrode (which we call the trap) they can tunnel across the barrier in a relatively short time (a few microseconds) compared to the loss time (a few hundred microseconds). This process of quasiparticle trapping allows us to make an STJ detector with a large absorber, good for photon collection, coupled to a small tunnel junction, good for fast, efficient quasiparticle readout.

Taking advantage of quasiparticle trapping, we can also make imaging STJ detectors based on charge division. To make a detector with one dimensional imaging, we start with a single Ta strip. We contact two opposite sides of the absorber with a pair of tunnel junctions (Fig 1.1). When a photon is absorbed in the strip, the created quasiparticles diffuse until they reach the two Al traps. The quasiparticles scatter down in each trap. They then tunnel and produce an excess current in each junction, allowing us to collect two charges, Q_1 and Q_2. The sum of the charges is again Q_γ, which is proportional to the photon energy. In addition, the ratio of the charges tells us where the photon landed along the absorber. If the photon lands in the middle of the absorber, the charge divides equally between the two junctions on average. If the photon lands close to one junction, almost all of the charge will be trapped in the nearby junction and very little will be collected by the far junction. In an ideal detector, the charge collected in each junction varies linearly with the absorption location [Krauss 1989].

The division of the charge between the two junctions is a statistical process, with each quasiparticle following a random walk. Thus, repeated measurements will produce a distribution of absorption locations, even if all the photons land in exactly the same place. The width of the distribution, ΔL, defines the spatial resolution of our detector. Besides the statistics of diffusion, ΔL can also be increased by additive noise in the tunnel junction and the electronic readout. If we interpret ΔL as the effective pixel size of the detector, we can then define the number of pixels as:

$$N_{pixels} \equiv \frac{L}{\Delta L} \tag{1.3}$$

where L is the length of the absorber. In general, N_{pixels} will be of the same order as the energy resolving power, R.

There is an intrinsic process in superconducting tunnel junctions, known as backtunneling, that we can exploit to give the junction charge gain. We do this by adding a Ta "plug" interrupting the wiring on the counter electrode side of

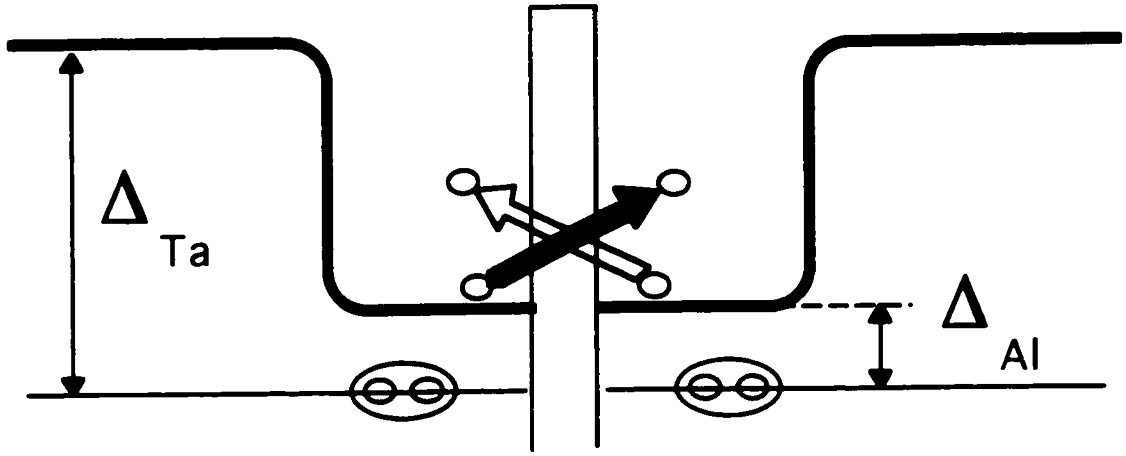


Figure 1.2 Energy band diagram of one junction of a backtunneling device in a modified excitation representation. Quasiparticles are shown as gray circles, suggesting their mixed electron-hole character. Tunneling is shown as diagonal transitions across the barrier, indicating that quasiparticles gain (lose) energy as they are accelerated (decelerated) by the bias voltage. Electron tunneling is indicated by black arrows. Hole tunneling is indicated by white arrows. Quasiparticles are confined in both Al electrodes by high gap Ta. At high bias voltage, only electron tunnel is allowed from left to right and only hole tunneling is allowed from right to left. This hole process is known as backtunneling. The two processes allow a single quasiparticles to circulate, tunneling multiple times. This gives the junction charge gain because both processes transfer a charge in the same direction.

the barrier (refer to Fig. 1.1). Fig. 1.2 is energy band diagram of one junction in a modified excitation representation. We see that the high gaps of the Ta absorber and plug confine excited quasiparticles near the tunnel barrier. Each quasiparticle is a coherent superposition of electron and hole, but it must tunnel as either a pure electron or pure hole. At sufficiently high bias voltages, a quasiparticle can only tunnel from left to right as an electron, gaining energy eV. A quasiparticle cannot tunnel from left to right as a hole, because it would lose energy eV and tunnel into the gap on the right side. After tunneling to the counter electrode, the quasiparticle can only backtunnel from right to left as a hole. Keeping both processes in mind, we see that a confined quasiparticle can circulate, first tunneling and then backtunneling. This cycle continues until the quasiparticle is lost to recombination. Because both tunneling and backtunneling transfer a charge in the same direction, this effect gives the junction a charge gain, p, equal to the average number of times a quasiparticle tunnels. Thus, we measure an integrated charge many times greater than the number of quasiparticles initially created.

Backtunneling gives the junction charge, but it also adds noise. Both tunneling and backtunneling are random processes, so the number of times different quasiparticles tunnel randomly varies. Thus, the gain associated with backtunneling varies from photon to photon, broadening the resolution. For a symmetric tunnel junction, the energy width due to backtunneling is

$$\Delta E_{Back} = 2.355\sqrt{\left(1+\frac{1}{p}\right)\varepsilon E_{\gamma}} \qquad (1.4)$$

where p is average number of times a quasiparticle tunnels **[Goldie 1994]**. This noise is minimized in the limit of a large gain. Still, even in the best case it contributes a width more than a factor of 2 greater than the Fano limit.

1.3 Previous and Concurrent Work

As we can see, there are a number of physical processes involved in the operation of an STJ detector. In this introduction, we have discussed the idealized aspects of these processes. The rest of this dissertation is largely concerned with answering the question, "How close does a real STJ come to this ideal behavior?" Many aspects of this question have been answered in previous work, much of it done by prior students here at Yale. The present work on STJ detectors for optical/UV application grew out of work to develop STJ detectors for X-ray astronomy.

Michael Gaidis [Gaidis 1994] began the work at Yale. He developed the geometry and basic theory of operation. He developed the successful fabrication process, including basic material science, that we use today. In fact, actual devices made by Gaidis were tested repeatedly through 1999 and they still worked when we stopped testing them. Gaidis made the first successful measurements of single junction STJ detectors at Yale. His thesis reports an energy resolution of 190 eV FWHM for 6 keV X-rays. (We can use as a bench mark for energy resolution 160 eV at 6 keV achieved by a standard high-performance, commerial semiconductor detector, the Amptek XR-100CR.) Stephan Friedrich continued the development of X-ray detectors [Friedrich 1997a]. He made the first measurements of double junction imaging detectors. He also developed the active voltage bias amplifier that we still use. His work improved the energy resolution of the detectors to 54 eV at 6 keV. The improved electronic readout developed by Friedrich opened the door to understanding much of the underlying physics of the devices. Kenneth Segall continued the work on X-ray detectors. Most of his work was done on the same physical devices as Friedrich, but further refinements of the electronics and measurement setup improved the energy resolution to 26 eV at 6 keV. He also developed a detailed microscopic model of the detectors, including extensive computer simulations. He used those models to complete much of our physical understanding of the device operation. In particular, he concluded that non-equilibrium processes in the detectors were limiting the energy resolution. With this understanding, he was able to quantitatively explain the energy resolution of the detectors and make predictions about how design changes could improve the resolution.

The development of optical/UV detectors started concurrent with Segall's work. The basic operational principle is the same as in the X-ray energy range. We modified and scaled device designs appropriately for lower incident energy. We developed an optical test system and fabricated new devices. We used the same basic electronics for the optical/UV measurements, although we improved the noise performance. We have also put much work into improving the repeatability and efficiency of the measurements system by refining the grounding and shielding of the system and the basic measurement protocol.

Development of X-ray detectors has also continued here at Yale. This work has been carried out by Liqun Li, Luigi Frunzio and the author. Recently fabricated X-ray devices have achieved an energy resolution of 13 eV at 6 keV, as good as any STJ detector in the world [Li 2001]. A next generation of devices have been designed and fabricated based on the predictions of Segall. Tests of these latest devices are ongoing.

Two other groups have made successful optical/UV single photon spectrometers. A group at the European Space Administration (ESA) produced the first results with any type of cryogenic detector [Peacock 1996]. They use STJs with a vertical geometry, meaning that the Ta layers and the Al junction electrodes are stacked on top of each other. (By contrast, we refer to our device geometry as a lateral geometry.) In the laboratory, they have measured an energy resolution of 0.25 eV at 5 eV [Verhoeve 1997] for a fiber-coupled STJ. In addition, they have also made several measurements using a lens-coupled 36 pixel array of STJs at the Walter-Hershel Telescope in the Canary islands [Verhoeve 2002]. The resolution of the lens coupled array is significantly degraded by the flux of infrared photons from the warm lenses. A group at Stanford University has also made optical/UV spectrometers based on TESs. In the laboratory, they have achieved an energy resolution of 0.15 eV at 5 eV for a fiber-coupled TES [Cabrera 1998]. They have also made astronomical measurements using a fiber-coupled TES [Cabrera 2002].

Chapter 2: Experimental Apparatus and Conditions

2.1 Device Fabrication

All of the devices discussed in this thesis were fabricated in the Yale Microelectronics Fabrication Center. Devices are fabricated in a six layer process in a high vacuum chamber with a typical base pressure of 1.5×10^{-7} Torr. The chamber is pumped by a pair of 8″cryopumps. We start with a thermally oxidized 2″ silicon wafer. The wafer is cleaned and then baked overnight in the vacuum chamber using a quartz lamp at a temperature of 350° C. The first deposition is the Ta layer. Immediately before deposition, the bare wafer is cleaned with an in-situ ion-beam gun to help ensure good film adhesion. The Ta is dc-sputtered at 960 W with an argon pressure of 6 mTorr . During the deposition the substrate is heated to 750° C with the quartz lamp. The Ta must be deposited hot so that it nucleates in the desired crystal phase [Face 1987]. The typical Ta thickness is 600 nm. For device runs OPS-F99 and OPS-E00, we sputtered titanium (Ti) into the vacuum chamber for approximately 30 minutes before the Ta deposition to reduce the background pressure. The typical pressure after sputtering Ti was 8×10^{-8} Torr. The Ta film is patterned using a positive photoresist process and wet chemical etching.

The second layer is the Nb contact. The Nb is patterned using a lift-off process, so an invertible photoresist is spun over the patterned Ta. The resist is patterned with a negative image and then the wafer is inserted into the vacuum chamber. Immediately before depositing the Nb, the Ta is cleaned with the ion-beam gun through the photoresist mask. This ion-beam cleaning removes any Ta-oxide or other contamination from the exposed Ta ensuring a good metallic interface between the Ta and Nb. The Nb is dc-sputtered at 330 W with an argon pressure of 11 mTorr. The typical film thickness is 200 nm. The wafer is then removed from the vacuum chamber and the film is lifted off in acetone.

The next deposition step is the Al trilayer. The wafer with the patterned Ta and Nb layers is cleaned and inserted into the vacuum chamber. For samples OPS-F99 and OPS-E00, we again sputtered Ti to lower the base pressure. While the Al evaporation source is preheated, the sample is ion-beam cleaned, again to ensure a good metal-metal interface between the Ta and Al trap. This is by far the most important interface. The Al trap layer is thermally evaporated at the high rate of about 20 nm/s. We evaporate at a high rate to reduce the exposure of the film to background gas and to promote the nucleation of large Al grains. The typical trap film thickness is 200 nm. The Al is evaporated from a pair of tungsten coils wrapped with Al wire. The parallel coils form an extended source that helps ensure a uniform deposition across the wafer despite the short working distance needed to achieve such a high deposition rate. After the deposition, we allow the sample to cool for a few minutes before isolating the chamber from the cryopumps and filling it with ultra-high purity oxygen gas to a pressure of 500 mTorr. We allow the Al film to oxidize for 120 minutes. At this point, we remove the oxygen by directly opening the cryopump valves. We then thermally evaporate the counter electrode Al film at a rate of 2 nm/s. We use a much lower deposition rate for the counter electrode film to avoid damaging the Al-oxide barrier. The typical counter electrode thickness is 75 nm.

The Al/Al-oxide/Al trilayer is patterned in a two-step, wet-etch process. First, we pattern positive photoresist in the shape of the trap layer and the entire trilayer is chemically etched in this shape. We then clean off the photoresist, and pattern a new layer of positive photoresist in the shape of the counter electrode, which is given a 2 μm separation from the edge of the trap. We then do a timed etch to define the counter electrode without etching the trap layer any further.

The next layer is an insulating layer with a via (a hole) that allows for contact to the counter electrode. We use SiO as an insulator and pattern it using a lift-off process. To do this, we spin invertible resist over the sample and pattern it. We place the wafer in the vacuum chamber and thermally evaporate Si0 at a rate of 2 nm/s. The typical thickness is 250-300 nm, enough to ensure step coverage over the entire trilayer. The sample is removed and the SiO is lifted off in acetone, leaving the via through the insulator to the counter electrode. The insulating layer does not cover the absorber in the optical/UV devices.

The final step is to deposit Al wiring. The Al wiring includes macroscopic pads for Pogo Pin contact to the samples and it contacts both the Nb ground lead and Al counter electrode. We spin invertible resist over the wafer and pattern it. The wafer is inserted into the vacuum chamber. The sample is ion-beam cleaned and Al is thermally evaporated at a rate of 2 nm/s. The total thickness is typically 250-300 nm, enough to ensure step coverage of the insulator. The sample is removed and the Al wiring is lifted off. At this point, the process is complete and the wafer is diced into individual samples.

2.2 Cryogenics

Our devices are tested in a custom-made, two-stage ^{3}He dewar. The dewar is comprised of a vacuum container, many layers of NASA super insulation, a liquid N_2 tank and shield, a liquid ^{4}He tank and shield, and two closed cycle ^{3}He pots. The stages are nested, with the ^{4}He stage inside the N_2 stage, and the two ^{3}He pots inside the ^{4}He stage. The second ^{3}He stage is not inside the first ^{3}He stage, but all supports and pumping lines for the second stage go through the first ^{3}He stage. The gaseous ^{3}He is stored in an attached tank at room temperature. Each ^{3}He pot has an individual charcoal pump contained inside the ^{4}He stage. The dewar is relatively compact compared to standard top-loading dewars. To gain access to the sample stage, the dewar is inverted onto a stand. The vacuum container, super insulation, N_2 shield and ^{4}He shield are removed in sequence. The sample space is large and has a copper cold finger that extends from the second ^{3}He pot.

The minimum time to reach the base temperature of 210 mK from room temperature is about 8 hours, although we usually run the cooling procedure overnight. We start by pumping out the vacuum container using a diffusion pump with a liquid N_2 cold trap. After pumping the dewar, we fill both the N_2 and ^{4}He tanks with liquid N_2 to pre-cool the dewar. The maximum pressure allowed before starting to pre-cool the dewar is 2 x 10^{-5} Torr. At this point, we usually let the dewar pre-cool overnight while continuing to pump, although the cooling process only takes about 2 hours. In the morning, we valve off the vacuum container once the pressure drops below the maximum allowed value of 2 x 10^{-6} Torr. At this point, we remove the liquid N_2 from the ^{4}He tank and we transfer liquid ^{4}He into the tank. The tank holds approximately five liters of ^{4}He. After allowing the ^{4}He stage to cool to 4 K, we top off the liquid ^{4}He and begin to pump on the bath. The entire bath is pumped using a Roots blower backed by a rotary pump. The maximum pumping speed of the Roots blower is 300 liters/s. The Roots blower is left on for the rest of the cryogenic run. Once the ^{4}He bath reaches its base temperature of 1.5 K, we begin condensing ^{3}He into the two pots by opening valves connecting the room temperature ^{3}He cylinder to the pots. After condensing for about an hour, we begin pumping on both ^{3}He pots by opening thermal switches connecting the pots to the ^{4}He bath and closing switches connecting the two charcoal pumps. After about two hours of pumping on the ^{3}He pots, the first pot reaches a temperature of 0.3 K and the second pot reaches the base temperature of 0.21 K. Often we refer to the two pots as the 0.3 K and 0.2 K stages. The ^{4}He bath typically limits the hold time of the dewar. The ^{4}He bath typically lasts for twelve hours after topping it off, which implies 8-9 hours of experimental time at base temperature.

2.3 Optics and Light Sources

We have developed an optical system for testing the devices. We start with two different light sources at room temperature. One light source is a mercury arc lamp, designed for calibrating spectrometers. The lamp's output spectrum contains a number of narrow spectral lines corresponding to the atomic transitions of the mercury vapor. The lines span the entire optical energy range and extend into the UV. The lamp has a total output of less than 5 mW, spread across the various lines. We use this lamp to provide single photons at varying energies. The second light source is a pulsed N_2 laser. It emits extremely bright pulses of UV light ($\lambda = 337$ nm, $E_\gamma = 3.68$ eV), lasting about 4 ns. This source can simulate high energy photons by illuminating the device with multiple UV photons that are absorbed on a timescale short compared to the device readout. Both of these light sources are fiber coupled. We can only use one source at a time, although it is straightforward to switch between the two.

The light from each source is collected by a fiber-optic patchcord and brought to a small optical bench. We use fused silica fibers that transmit light of wavelength down to 200 nm. At the entrance of the optical bench, the end of the fiber is placed at the focus of a miniature collimating lens that launches the light radiating from the fiber onto the optical bench. The optical bench has two types of filters. Metallic neutral density (ND) filters are the first type. These filters attenuate light uniformly across a broad spectrum, specifically from 200 nm – 2000 nm. We use them to adjust the intensity of the light that reaches the devices. The filters are made of fused silica, so that they transmit UV light, and they attenuate the light using varying thicknesses of a reflective metallic coating. The filters are characterized by a neutral density value, ND, which is the logarithm (base 10) of the attenuation. We have twelve ND filters placed in two filter wheels. The filter wheels allow us to rotate different filters into and out of the beam line without having to realign the optics. One wheel has filters with neutral density values of ND = {0.03, 0.5, 1.0, 1.5, 2.0, 2.5}. The other wheel has values of ND = {0.03, 0.1, 0.2, 0.3, 0.4, 0.5}. The two wheels are placed in series and together they allow us to vary the attenuation over 3 orders of magnitude in logarithmic steps of 0.1. We use a third filter wheel to hold two filters with values of ND = {1.5, 3.0} that allow us to further increase the attenuation. We use these ND filters to adjust the intensity of both the laser light and the light from the mercury lamp. The output of the laser is also attenuated by a high-power, fiber-coupled filter with a value of ND=4.0 before reaching the optical bench.

Interference filters are the second type of filter. These are narrow bandpass filters, with a typical FWHM of only 10 nm. The filters have two highly reflective metal surfaces, separated by a precise distance. The two surfaces reflect all wavelengths strongly, attenuating the light. However, they also create a resonant cavity, defining the center wavelength of the filter. Light of the resonant wavelength builds in the cavity to a relatively high intensity, such that despite the fact that most of it is reflected by the far surface, the total transmission is of order unity at the center wavelength. We have four interference filters, each with a center wavelength matched to one of the spectral lines of the mercury lamp. These filters are placed in a filter wheel that allows us to easily change which mercury line is shone on the device. We have filters that select the 254 nm (4.89 eV) UV line, the 436 nm (2.85 eV) violet line, the 546 nm (2.27 eV) green line, and the 690 nm (1.80 eV) red line. The peak transmission varies such that the intensities of the all lines are the same after the filters. (Actually, the red filter is not exactly matched to the red mercury line. It is a generic filter with a center wavelength close to the red line.) These filters are not used with the laser, which naturally emits a single wavelength.

After being conditioned by the various filters on the optical bench, the light is collected by a second, identical collimating lens that refocuses the light back into a fiber splitter. The light collected by the input fiber of the splitter is equally split between two output fibers. One of the output fibers goes to the dewar. The other one goes to a room temperature photomultiplier tube (PMT) that is used to monitor the intensity of the light sent into the dewar. The PMT is housed in a custom-made metal case that shields it from stray light. The fiber is connected to a coupler that screws directly into the PMT case.

The optical bench was designed with a collimating lens on both ends to minimize the amount of stray light focused into the fiber splitter. The second lens only focuses light into the splitter that is moving parallel to the optical axis of the bench. Thus, the solid angle from which the lens collects light is relatively small. In fact, if we cover the optical bench with an ordinary cardboard box with slits cut for the fibers to enter and exit, the amount of stray light entering the fiber is undetectable by the PMT, even with the room lights on. By undetectable, we mean that the PMT current generated by the stray light is negligible compared to the thermal dark current of the PMT, which is 10 nA. The typical intensities we use to test STJs correspond to PMT currents of 1-3 μA. Thus, the ratio of source light intensity to stray light intensity is $\sim 10^3$ or better.

One end of the fiber splitter brings the light to the dewar. The light is coupled through the bulkhead of an rf-shielded enclosure (which we will discuss later) using a fiber coupler. Inside the shielded enclosure, a final fiber patch cord brings the light to a custom made vacuum fiber feedthrough mounted to the dewar. This feedthrough was made starting with commercial Quick Connect parts. Quick Connect parts are designed to bring tubes or conduits into a vacuum system. They have a stem that accepts a tube and a collar that can be screwed down to compress an o-ring against the sides of the tube. We welded a stem onto a custom flange made to fit the dewar. We then machined a brass Quick Connect blank cap to make the feedthrough. The blank cap is designed to plug the stem when it is not being used. We show a schematic of the modified cap in Fig. 2.1. We first dimpled the top of the cap. We then drilled a large hole down the axis of the cap from the bottom leaving only a small thickness of brass. We then drilled a short, 600 μm diameter hole connecting the top and bottom. We then fed an Al-coated UV fiber, with a total diameter of about 500 μm, through the hole. Finally, the gap between the fiber and the cap was sealed by adding a small amount of Crazy glue to the dimple on the top of the cap. We have successfully tested the feedthrough down to a vacuum pressure of 5×10^{-8} Torr, which is the limit of our diffusion pump system.

Inside the dewar, the Al-coated fiber is first wrapped around the liquid N_2 can and then wrapped around the liquid ^{4}He can before entering the sample space. A length of about 1 meter of the fiber is fixed to each can with Al tape. This is done to progressively cool the fiber. The fused silica fiber we use is *not* treated to remove water and so, water impurities in the fiber absorb infrared

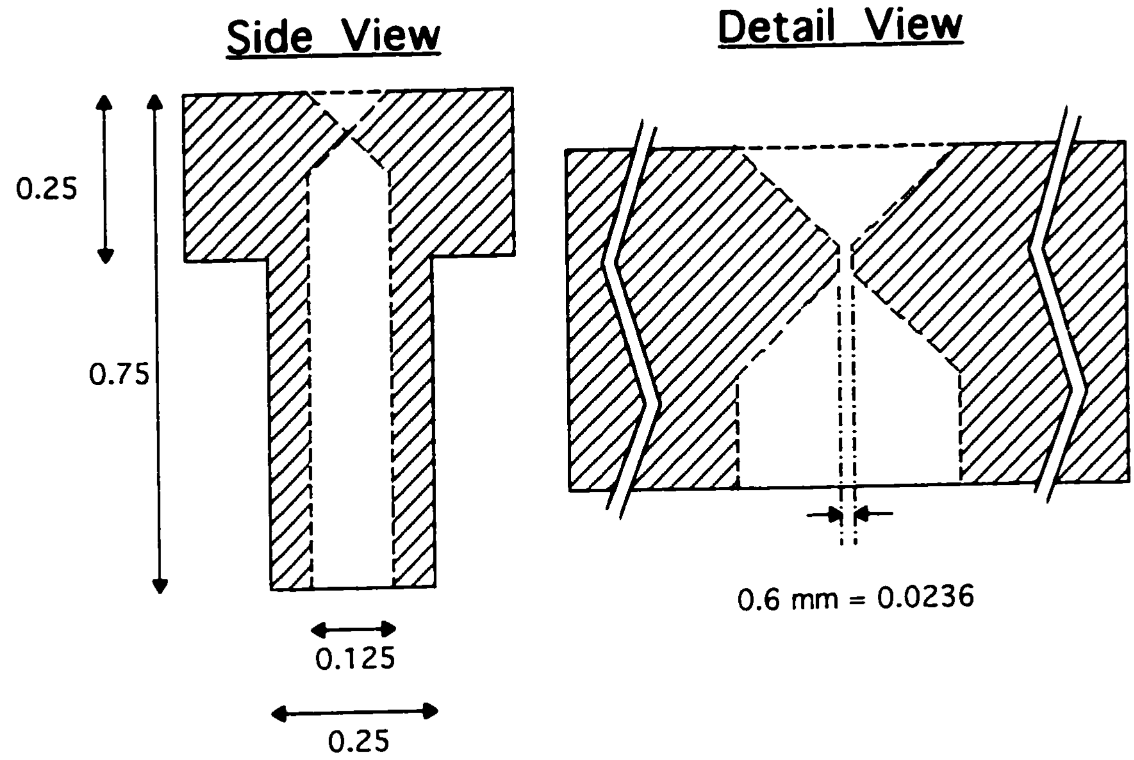


Figure 2.1 Schematic of fiber feedthrough made from Quick Connect blank cap. Measurements are in inches, unless otherwise noted.

radiation. Thus, the progressively cooled fiber acts as an infrared filter, each cooled section absorbing blackbody radiation from the previous, warmer section.

We use an Al-coated fiber for a number of reasons. First, it allows for a good vacuum seal at the feedthrough. Second, it improves the transmission of the fiber in the deep UV range. The Al coating does this by reflecting light that escapes the dielectric back into the fiber. The coating also greatly improves the mechanical strength of the fiber, allowing it to survive many thermal cycles. Finally, the Al coating may improve the thermal contact of the fiber with the dewar.

At the sample stage, the fiber is suspended from the liquid ^{4}He stage and the end of the fiber is brought within a few millimeters of the device on the 0.2 K cold finger. The fiber does not physically touch the 0.2 K stage anywhere. The fiber is roughly aligned by eye to the device. To do this, a piece of lens paper is placed on top of the device. A laser pointer is then shone into the fiber at the vacuum feedthrough. The laser light coming through the fiber produces a spot about one millimeter in diameter that is aligned to the device. The lens paper is thin enough to allow us to see the wiring of the devices underneath but opaque enough to make the laser spot visible. We estimate that only about one part in 10^4 of the light exiting the fiber lands on the device. While this inefficient coupling would clearly be inadequate for a real spectroscopy system, it is adequate for testing the devices.

Besides using the optical fiber to test the devices, we can also use it to warm the detector tunnel junctions above their transition temperatures for a short period of time during a cryogenic run. While this was not a design goal of the optical system, it has proven to be an extremely valuable addition to the experimental setup. (In fact, when testing X-ray devices, we still position the optical fiber to allow the junctions to be heated.) A variety of problems, including the presence of stray magnetic fields during cool down or an oscillating amplifier, can lead to fluxons becoming trapped in the electrodes of the detector tunnel junction. The presence of this trapped flux can degrade the performance of the detector in a number of ways, from instability in the bias to decreased dynamic resistance. It is very difficult to remove the fluxons as long as the junctions remain superconducting and, in the past, trapped flux ended many cryogenic runs. The only reliable way to remove the trapped flux is to warm the junctions above their transition temperature. Warming the entire cold stage above the transition temperature of Al (~ 1.3 K) is impractical. In addition, trying to locally heat the immediate surroundings of the device electrically is difficult because the copper cold stage conducts heat well. With the optical fiber, however, we can illuminate just the junctions and the surrounding surface of the substrate with an intense burst of light (from a laser pointer) and warm them above the transition temperature for a few seconds without warming the rest of the cold stage. We see no substantial decrease in run time due to this process. This laser cycling works reliably and has greatly improved the efficiency with which we can test detectors.

2.4 Electronics

The electronics used to read out the signal from our detectors has been discussed extensively in previous dissertations [Segall 2000], [Friedrich 1997a]. I will only provide a general description, emphasizing changes that have been made. We start with a commercial amplifier, the Amptek A250. One unique property of the A250 is that we can couple it to a discrete input FET, chosen to best match the characteristics of our detectors. We use the 2SK146 FET, which is two 2SK147 FETs packaged together in one metal can. We use the amplifier in a transimpedance configuration with a resistor providing feedback from the output of the A250 to the gate of the 2SK146. The total amplifier has a low input impedance (~ 100 Ω) and draws current from the detector across the feedback resistor producing a voltage at the output. We typically use a 1 MΩ feedback resistor.

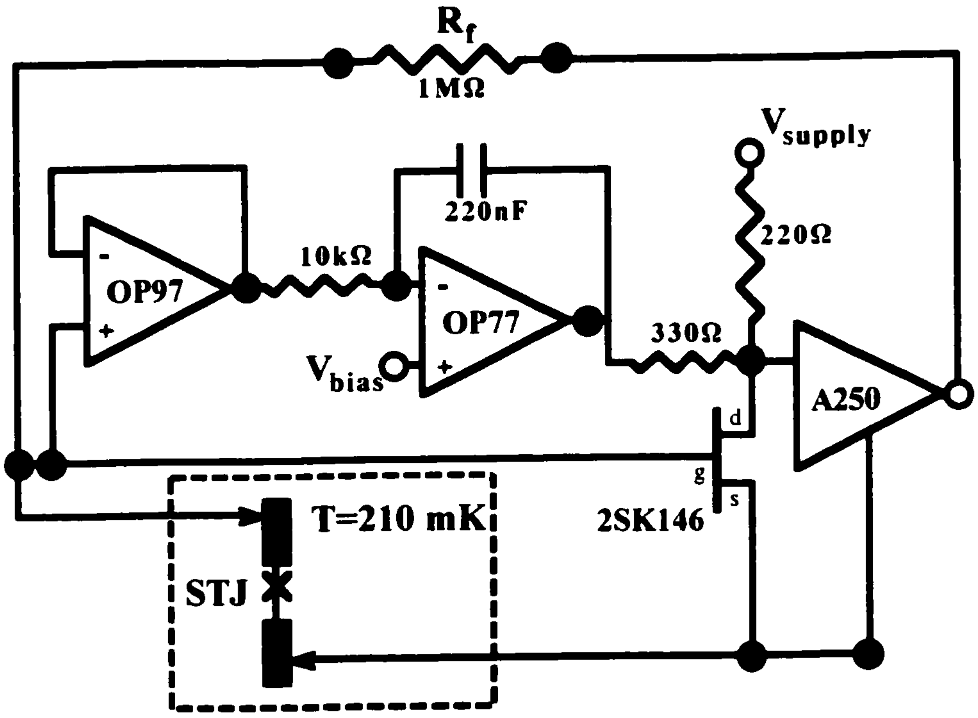


Figure 2.2 Simplified schematic of the signal preamplifier.

The A250 was designed to be used as an AC coupled amplifier. We add extra circuitry in parallel with the input FET that provides an active DC voltage bias for our detectors. The extra circuitry consists of an OP97 op-amp configured as a voltage follower connected to an OP77 op-amp configured as an integrator. Referring to Fig. 2.2, we see that the input of the OP97 is connected to the device (along with the gate of the input 2SK146) and monitors the voltage across the device. The OP77 compares that voltage to a reference voltage and injects current (through a resistor) into the input of the A250. The OP77 is configured as an integrator so that it provides a large gain at DC, but contributes little noise (or gain) at signal frequencies. Both the OP97 and OP77 are precision op-amps with small voltage offsets and drifts.

The most significant improvement made to the electronics was to cool the feedback resistor. For X-ray measurements, the Johnson current noise of the 1 MΩ feedback resistor, 130 fA/√Hz, dominates the electronic noise of the measurement setup. However, the electronic noise only contributed a small amount of broadening to the total energy width. Since the signals are much smaller in the optical/UV energy range, however, the electronic noise needed to be reduced. This was done by moving the feedback resistor from room temperature to the ^{4}He stage, so that its temperature is 1.5 K during measurements. This should reduce the Johnson noise of the feedback resistor from 130 fA/√Hz to 9 fA/√Hz. Along with connecting the amplifier to the cold feedback resistor, we needed to provide a room temperature feedback path so that we could turn on and test the amplifier without having it connected to the detector. At first we tried putting a warm 100 MΩ feedback resistor in parallel with the cold resistor. The warm 100 MΩ resistor provided a feedback path while not contributing a significant amount of Johnson noise. However, when we would first connect the cold 1 MΩ resistor in parallel, the sudden change in the feedback condition of the amplifier produced a large transient response that tended to trap flux in the detector tunnel junction. In the end, the solution we found was to use two room temperature switches, labeled S1 and S2 in Fig. 2.3.

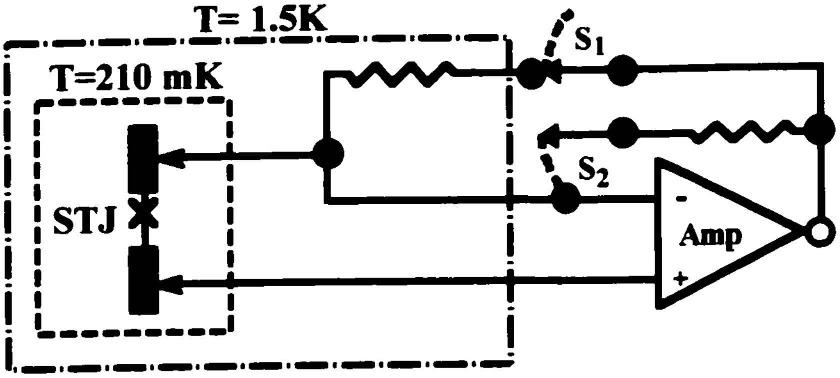


Figure 2.3 Simplified schematic of AC electronic circuit, showing switches S_1 and S_2 that allow us to switch between the warm and cold feedback resistor.

These switches allow us to first bias the junction with a warm feedback resistor (S1 open, S2 closed), then connect the cold feedback resistor in parallel using S1, and finally disconnect the warm resistor using S2.

We acquire data using a digitizing oscilloscope, the Nicolet Integra 40. The oscilloscope has four differential input channels. It digitizes the input to each channel with 12 bit precision and a maximum sampling frequency of 20 MHz. Each channel of the oscilloscope has a one million sample memory. When the oscilloscope memory is full, data is downloaded over a GPIB bus to a personal computer. We use the oscilloscope to acquire both photon-induced current pulses and noise traces for diagnostics. When we acquire current pulses, many pulses are saved in the oscilloscope memory before being downloaded. When acquiring noise traces, the memory is typically filled with a single long trace. Once downloaded, the noise traces are fast Fourier transformed to produce noise spectra. Then a number of spectra (typically 100) are averaged, and the average spectrum is saved to disk. The current pulses are saved to disk as is.

2.5 Grounding, Shielding and Filtering

The smaller signals associated with optical/UV photons necessitated reducing electronic noise in the measurement system, including contributions from electromagnetic interference. Thus, a large amount of work has gone into improving the grounding, shielding and filtering in the measurement system. While the development of the present configuration has been guided by some general principles of grounding and shielding, much has also been decided through trial and error.

When designing or analyzing a design for grounding and shielding a measurement system, it is important to distinguish between the earth ground of the shield and the electrical ground of the measurement circuit. The earth ground of the shield literally connects the shield to the Earth, which is usually the zero potential of power systems and radio transmission. In general, no currents associated with the signal to be measured should flow to earth ground and conversely, currents flowing to earth ground should not be detected by the measurement circuit. The electrical ground of the measurement circuit is the node defined to be zero potential in the circuit. Although this node is generally treated as special, it is just another set of conductors. Currents or voltages induced in the electrical ground by interference will in general be detected by the measurement circuit, becoming noise. We generally do connect the electrical ground of the circuit to the shield and earth ground at one or more points. However, this forces them to be at the same potential only at DC. The potential on the grounds can vary spatially on size scales down to the wavelength of the electromagnetic interference. This implies that radio frequency interference can create different potentials in the grounds even if they are connected together. In addition, when designing a system for low-frequency shielding, against 60 Hz interference for instance, it is important to consider the path of currents generated by an induced electromotive force (EMF). A successful design forces induced currents to flow in the conductors of the shield and not the electrical ground, where the currents become a source of noise.

The starting point of our design is a commerial shielded enclosure made by Lindgren RF-Enclosures, Inc.. The shielded enclosure is a cube, 4 feet on a side. Each side is made from two 1/4" steel plates separated by a couple inches of plywood. There are two doors, one on the side and one on the top, that allow

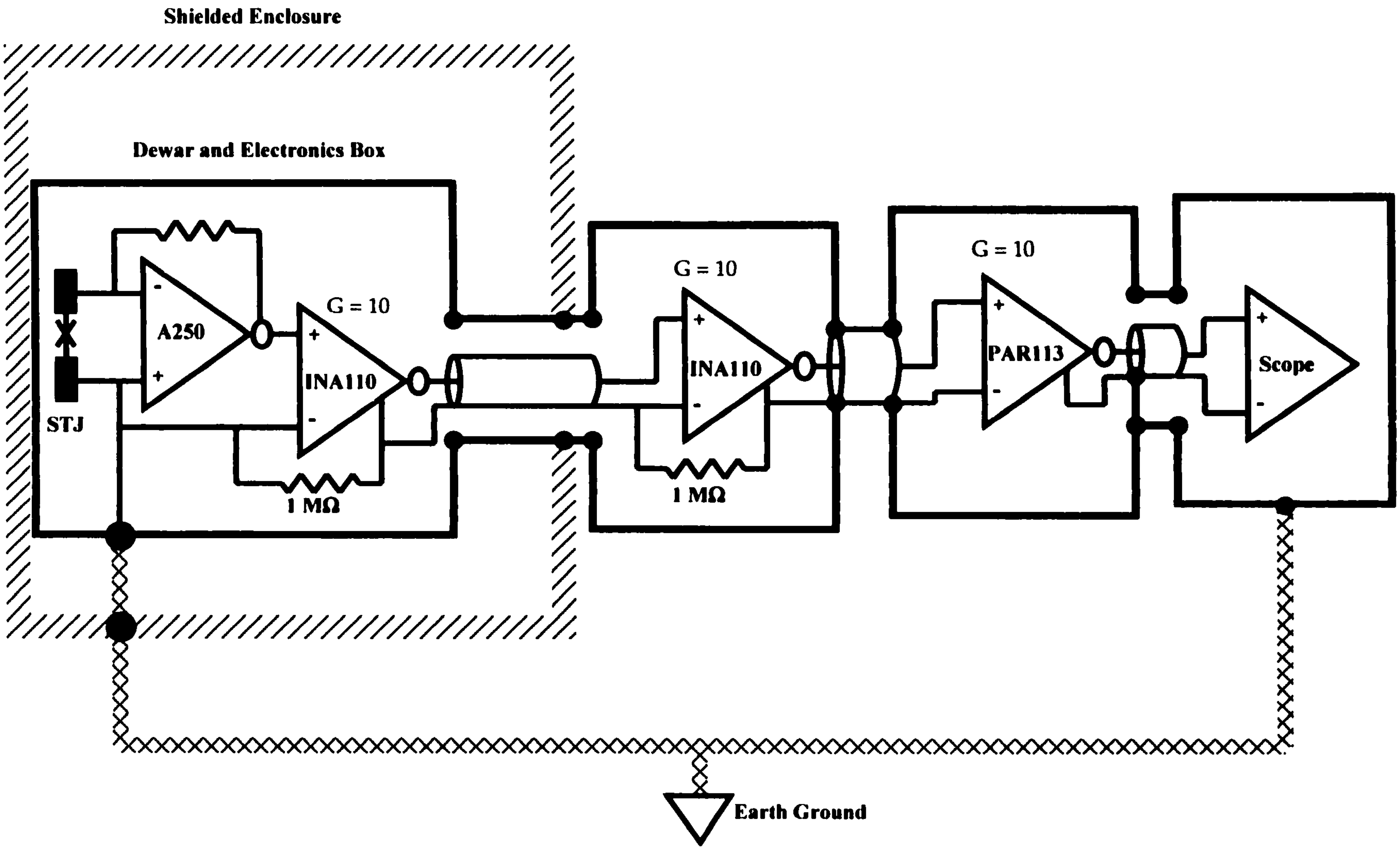


Figure 2.4 Schematic of the grounding and shielding design. Each of the electronics boxes has a separate battery supply, except the oscilloscope which uses line power. The A250 and INA110 in the first box have two separate battery supplies and grounds. Each of the INA110s and the PAR113 has a voltage gain of 10.

access to the inside of the enclosure including inserting and removing the dewar. The doors close tightly with a set of copper "fingers" welded to each of the doors' steel plates. There is a 1" tube in one wall that allows us to pump on the dewar while it is inside the enclosure. The tube is filled with cooper wool to allow gas flow while preventing electromagnetic radiation from entering. The enclosure also has a power line filter that allows filtered, AC power to be brought into the enclosure, although we have not used AC power in the enclosure to date. There is a 1/2" diameter brass screw threaded through one side of the enclosure. The screw contacts both the inner and outer steel plates, electrically connecting them at one point. We connect this screw to a copper cold-water pipe using a 1" thick braided grounding strap. This point serves as the earth ground of the shield. Connecting a DC ohmmeter between the shielded enclosure and the third prong of an electrical outlet, we measure a resistance of about 1 Ω. (Earthing the enclosure is also an electrical safety issue.)

Inside the enclosure, the dewar is connected by a grounding strap to the brass screw, earthing the dewar. The electronics are enclosed in a metal box which is connected to the dewar using a military connector. The dewar and electronics box form a continuous metal shield. All of the electronics inside the enclosure are battery powered. The detector, contained inside the dewar, is connected to room temperature differentially using a twisted pair of wires. The negative signal wire, or electrical ground, is connected to a ground plane and then to the metal box at one point, physically close to the source of the AC amplifier's input FET. The electrical circuit itself has two distinct grounds: an input ground and an output ground. The input ground includes the negative signal wire coming from the detector to room temperature. The output ground is isolated from the input ground by a set of INA110 instrumentation amplifiers that prevent noise from coupling back to the input. There are separate battery supplies for electronics connected to the input ground and the output ground. In practice, the input and output grounds are connected by a 1 MΩ resistor, that prevents the two grounds from developing different electrostatic potentials. This reduces the isolation to some degree, but any induced EMF should drop across this resistor and not create signals in the two grounds. The output of the instrumentation amplifier and the output electrical ground are brought out of the box on a triax cable. The triax cable has three coaxial conductors: the center conductor, the inner ground and the outer shield in order from inside to outside. The output of the instrumentation amplifier is connected to the center conductor and the output ground is connected to the inner ground. The outer shield of the cable contacts the electronics box. The triax cable runs to the wall of the enclosure where a triax feedthrough brings the center conductor and inner ground outside of the box. The feedthrough goes through a metal plate screwed to the outer wall of the enclosure. There is hole in the inner wall behind the plate.

Outside of the enclosure, a short triax cable brings the center conductor and inner ground to another battery-powered electronics box. The outer shield of the triax connects the outer wall of the enclosure to the metal case of the electronics box. Inside the metal case, the center conductor and the inner ground are connected differentially to another instrumentation amplifier. This instrumentation amplifier amplifies the signal coming out of the enclosure with a gain of 10. The output of the amplifier is connected to the center conductor of a standard coax connector. The output ground of the amplifier is connected to the outer conductor of the coax, which is also connected to the metal case of the electronics box. The instrumentation amplifier provides a high degree of isolation between the electronic ground coming out of the enclosure and the shield, preventing noise from coupling back. At this point, our signal is carried by a standard coax cable for a few feet. This is the only place that a signal wire (in this case, the ground) is unshielded by another conductor. We go over from a triax cable to a coax cable to facilitate connecting other components and amplifiers. It does not compromise the shielding too much at this point because the signal is amplified enough by the last instrumentation amplifier that the level of interference is not significant.

The signal is now brought by the coax cable to a Princeton Applied Research (PAR) 113 amplifier, which is battery powered. The signal is amplified another factor of 10 and is also filtered to remove noise. The output of the PAR113 is then connected to the input of the digitizing oscilloscope, which records the signal. For many of the measurements in this dissertation, the PAR113 and oscilloscope were connected with a standard coax. This configuration connects the electronic ground of the amplifiers outside of the enclosure to the case of the oscilloscope, which is in turn connected to the earth ground of the power outlet. The electronic ground of the outside circuit is also connected to the outer wall of the shielded enclosure, which is earthed. Thus, this configuration creates a large ground loop that is completed by the two earth connections. However, by the time the signal reaches this point, it has been amplified by a factor of 1000. Consequently, in early measurements, the 60 Hz interference coupled by this ground loop was unimportant when referred to the input.

In later measurements, when the electronic noise had been reduced, 60 Hz interference dominated the low-frequency noise. The solution we found was to connect the output of the PAR113 to the oscilloscope differentially. The oscilloscope that we use can measure a differential signal by connecting the plus and minus signal wires as the center conductors of two standard coax cables. In our setup, the signal is carried from the PAR113 on a triax cable. At the oscilloscope an adaptor connects the center conductor and inner ground of the triax to the center pins of two coax connectors. The outer shield of the triax connects to the outer conductors of the coax connectors, which connect to case of the oscilloscope. In this configuration, there is still a short section of the electrical ground (between the second INA110 and the PAR113) that is part of the shield. However, we do not observe any 60 Hz interference, due to the high signal level at this point.

So far, we have described the AC signal path. There are also electronics that read out the DC bias voltage and current of the detector. The bias voltage and current are read differentially by a pair on INA110 instrumentation amplifiers. The outputs of the instrumentation amplifiers are then connected to a pair of optical isolation amplifiers that provide a very high degree of input-output isolation. The outputs of the isolation amplifiers and their output grounds are carried to the wall of the enclosure by standard coax cables. Outside of the enclosure, the pair of coax runs to the two inputs of an analog X-Y recorder. The bias current and bias voltage signals are read differentially by the X-Y recorder. The differential inputs of the X-Y recorder are very well isolated from the recorders case, which is connected to earth ground. We have never seen any 60 Hz or other interference in the AC signal related to the DC circuit being connected to the X-Y recorder.

All of the AC and DC lines coming out of the shield enclosure are filtered. The filters are custom-made T-filters that filter the signal coming out the enclosure, but also filter any interference or radiation coupling back into the enclosure. The filters are constructed in a way the prevents radio or microwave signals from coupling radiatively around the lumped elements that form the filters. To make the filters, we start with a metal box with BNC connectors on either end. A thick Al plate is inserted into the box, dividing the two sides. The inside surface of the box is prepared to ensure good electrical contact between the walls of the box and the divider. The edges of the divider are tapped with holes that allow screws to press the walls of the box firmly against the divider. A hole is also tapped through the divider along the axis of the BNC connectors. A commercial feedthrough capacitor is bolted into the hole. (We actually use a commercial LC pi filter, but the inductance is negligible at signal frequencies. Still, it may help suppress high-frequency interference. The feedthrough filter is made by Spectrum Control, Inc. and is part number 1250-003.) Finally, a resistor is soldered on each side, connecting the center conductor of each BNC connector to the center pin of the feedthrough capacitor. The filters on the AC lines have a cut-off frequency of 150 kHz. The filters on the DC lines have a cut-off frequency of 1 kHz.

Chapter 3

Physical Noise Processes: Thermodynamic Fluctuations

3.1 Introduction

There are myriad physical processes that add noise to our devices, meaning they increase the uncertainty in the energy of the incoming photon. In this chapter, we will discuss thermodynamic fluctuations of the number of steady-state quasiparticles in the Al electrodes of the STJ [Wilson 2001b]. We will show that these fluctuations lead to excess current noise in the tunnel junctions, which imposes an important limit on the resolution of backtunneling STJs.

Statistical mechanics elucidates the microscopic origin of the laws of thermodynamics. It connects the thermodynamic variables of a system to the ensemble averages of microscopic quantities. Despite the fact that it is only a theory of average quantities, thermodynamics is very successful at describing the behavior of macroscopic system. This fact can be understood if we consider the fluctuations of thermodynamic variables predicted by statistical mechanics. For example, consider a small container filled with a gas that can exchange energy and particles with a reservoir. This type of system is described by the grand canonical ensemble. In equilibrium, the chemical potential and the temperature of the gas in the container equal those of the reservoir and are fixed. However, the energy and number of particles in the gas are variables and can fluctuate. For instance, if the container holds an ideal gas with N^0 particles on average, the grand canonical ensemble tells us that the fluctuations are exactly $(N^0)^{1/2}$ [Pathria 1972]. If N^0 is macroscopic (of order Avogadro's number) then the relative fluctuations will be very small, of order one part in several trillion. The same is true for the fluctuations in the energy of the gas. The diminutive scale of these fluctuations implies that a macroscopic gas is described very well by its average, thermodynamic variables.

On the other hand, the same analysis suggests that fluctuations of thermodynamic quantities can be important in "small" systems. For instance, if the average number of particles in the gas is only about 10^4 then the relative fluctuations will be 1%. In this case, it is not clear that the average, thermodynamic variables can describe the system adequately. In any case, fluctuations with this relative size can have important, measurable consequences. To appreciate these consequences, it is crucial to understand not only the magnitude of the fluctuations but also the dynamics of equilibrium. When the gas is in equilibrium with the reservoir, the average flow of particles entering the gas from the reservoir cancels the average flow of particles leaving the gas. However, the flow of particles into and out of the gas is a random process, so that the two flows do not exactly cancel at all times. This causes the number of particles to fluctuate in time. Furthermore, if the characteristic time that a particle spends in the gas before returning to the reservoir is τ, then we expect the frequency spectrum of the fluctuations to be spread over a bandwidth $1/\tau$.

The thermal quasiparticles in the Al electrodes of an STJ detector form a Fermi gas. This quasiparticle gas is coupled to a particle and energy reservoir formed by the Cooper pair and phonon systems. In backtunneling devices, the quasiparticle gas in the Al electrodes is isolated by the high-gap Ta on both sides of the junction. In effect, the electrodes form an Al box confining the quasiparticle gas. In typical devices, the box has a volume, *Vol* = 100 μm^3. Particles flow into and out of the gas only through the processes of quasiparticle generation and recombination. Thermal quasiparticles in the Al cannot diffuse into the Ta because the difference between the superconducting energy gap of Ta (Δ_{Ta}= 700 μeV) and the energy gap of Al (Δ_{Al}=180 μeV) is much greater than $k_B T \sim$ 20-40 μeV. There are no thermal quasiparticles in the Ta at our base temperatures. Based on our general discussion above, we expect that the number of quasiparticles in the Al will fluctuate on the time scale of quasiparticle recombination.

We can calculate the quasiparticle number in thermal equilibrium by integrating the BCS density of states. Assuming that the quasiparticles obey a Fermi distribution and that $k_BT << \Delta_{Al}$, we find

$$N(T) = D(\varepsilon_F)Vol\sqrt{2\pi\Delta_{Al}k_BT}\exp\left(-\frac{\Delta_{Al}}{k_BT}\right) \tag{3.1.1}$$

where $D(\varepsilon_F)$ is the two-spin electron density of states at the Fermi energy. In our measurements $k_BT << \Delta_{Al}$, so the Fermi gas is nondegenerate. Its density is less than 10^{-3} that of excitations in normal-state Al at this temperature.

3.2 Theory: One-Variable Master Equation

We now develop a model that connects the frequency dependence of the fluctuations to the dynamics of quasiparticle generation and recombination. To treat fluctuations in the system, we construct a master equation similar to the Fokker-Planck equation. This differential equation describes the probability distribution of the occupancies of various subsystems (levels). We follow the treatment by van Vliet of generation-recombination noise in semiconductors [van Vliet 1965], except that we generalize the description to allow for transitions that involve an arbitrary number of particles, e.g., two quasiparticles recombining. The Al box is well described by the three level system of Rothwarf and Taylor [Rothwarf 1967]. The occupants of the levels are quasiparticles in the box, phonons with energy $E_\omega > 2\Delta$ in the box, and phonons with energy $E_\omega > 2\Delta$ in the bath. We can in fact predict the fluctuation of an arbitrary number of levels. However, the development is not particularly illuminating. We will instead present a detailed derivation for a simplified two-level system, and quote results for a multilevel system later.

We can consider one level of our system to be quasiparticles. The second level could be Cooper pairs or quasiparticles in traps or something else, depending on the exact nature of the system that we are trying to model. Regardless of exactly what the second level is, it is not in general independent of the first level because the total number of excitations in the two levels is constrained. For instance, the number of quasiparticles plus Cooper pairs is constrained by the total number of electrons, due to overall charge neutrality. Furthermore the creation of two quasiparticles implies the loss of one pair, and vica versa. Therefore, we only need to count the number of quasiparticles, N, and can describe our system with a one variable master equation:

$$\begin{aligned}\frac{\partial P(N,t|k,0)}{\partial t} = -[g(N)+r(N)]\cdot P(N,t|k,0) + g(N-\delta N)\cdot P(N-\delta N,t|k,0) \\ + r(N+\delta N)\cdot P(N+\delta N,t|k,0)\end{aligned} \tag{3.2.1}$$

where $P(N,t|k,0)$ is the probability that there are N quasiparticles at time t given there were k quasiparticles at t=0. The function g(N) is the probability per unit time that there will be a generation event in the box when there are N quasiparticles. In other words, g(N)dt is the probability of a generation event in the time interval dt. Similarly, the function r(N) describes the probability per unit time of recombination. The parameter δN is the number of quasiparticles added (removed) by a generation (recombination) event. We can understand the structure of the master equation quite simply. It describes the rate of change of the probability that there are N quasiparticles in the system. The rate of decrease in the probability equals the probability that there are N quasiparticles times the probability per unit time that there will be a generation or recombination event. This is what the first term in the master equation represents. The rate of increase in the probability is equal to the probability that the system is one generation event away from having N quasiparticles times the probability per unit time that there will be a generation event plus a similar term for recombination.

The master equation is an countably infinite set of coupled differential equations. Luckily, we do not need to solve the master equation for it to be useful. We can instead use the master equation to construct much simpler equations for quantities like the variance and correlation function of the fluctuations.

We begin by calculating the variance of the fluctuations. The variance is an equilibrium property, so we can set the left side of the master equation to zero. If we then multiply the equation by N, and sum over all possible N, we get the simple relationship:

$$\langle g(N)\rangle = \langle r(N)\rangle \tag{3.2.2}$$

where the angle brackets mean the expectation value over all N. If we expand both g(N) and r(N) in a Taylor expansion around the equilibrium value N^0, we get

$$g(N^0)+\tfrac{1}{2}g''(N^0)\langle\Delta N^2\rangle = r(N^0)+\tfrac{1}{2}r''(N^0)\langle\Delta N^2\rangle \tag{3.2.3}$$

where the primes indicate the derivative with respect to N and $\Delta N = N - N^0$. The first order terms vanish because $\langle\Delta N\rangle = 0$ in equilibrium. In addition, in most cases g(N),r(N) ~ N^2 and $\langle\Delta N^2\rangle$ ~ N, so we can neglect the second order terms and simplify to

$$g(N^0) \approx r(N^0). \tag{3.2.4}$$

This is the reasonable statement that the generation and recombination rates must balance in equilibrium.

If we again set the left-hand side of the master equation (3.2.1) to zero, multiply by N^2 and sum over all N, we get the relationship

$$\left\langle\left(N+\frac{\delta N}{2}\right)\cdot g(N)\right\rangle = \left\langle\left(N-\frac{\delta N}{2}\right)\cdot r(N)\right\rangle. \tag{3.2.5}$$

If we again expand g(N) and r(N) around N^0 and use (3.2.3) to simplify, we can find the following expression for the variance of the fluctuations

$$\langle\Delta N^2\rangle = \delta N\frac{r(N^0)}{r'(N^0)-g'(N^0)} \tag{3.2.6}$$

where we have again neglected second order terms in the final expression.

We can also use the master equation to calculate the power spectrum of the fluctuations. To do this, we first calculate the autocorrelation function of the fluctuations and then compute its Fourier transform. The autocorrelation function at lag u is defined as:

$$\Phi(u) = \langle N(0)N(u) \rangle = \sum_k \sum_j k \cdot j \cdot P(k,0;j,u) \tag{3.2.7}$$

where $P(k,0;j,u)$ is the *joint* probability that there are k quasiparticles at t=0 and that there are j quasiparticles at t=u. (By lag we mean the amount of time that one signal is shifted with respect to the other). We can simplify this expression by factoring the joint probability distribution into $P(k,0;j,u) = P(j,u|k,0) \cdot P(k,0)$ giving

$$\Phi(u) = \sum_k k \cdot P(k,0) \sum_j j \cdot P(j,u|k,0) = \sum_k k \cdot \langle N \rangle_k \cdot P(k,0) \tag{3.2.8}$$

where $P(j,u|k,0)$ is the conditional probability of having j quasiparticles at t=u given that there were k at t=0 and $\langle N \rangle_k$ is the expectation value of N given that there were k quasiparticles at t=0.

To further simplify this expression, we start by deriving a differential equation for $\langle N \rangle_k$ using the master equation. In this case, we need to use the full master equation (3.2.1) without setting the time derivative equal to zero. If we multiply both sides by N and sum over all N, we get the equation

$$\frac{d}{du}\langle N \rangle_k = \delta N(\langle g(N) \rangle - \langle r(N) \rangle). \tag{3.2.9}$$

We cannot solve this equation explicitly, because we do not know the expectation values on the right-hand side. However, we can find an approximate solution by again expanding g(N) and r(N) around N^0. We find the simple result

$$\frac{d}{du}\langle \Delta N \rangle_{k-N^0} = -\frac{\langle \Delta N \rangle_{k-N^0}}{\tau} \quad ; \quad \tau \equiv \frac{1}{\delta N} \frac{1}{r'(N^0) - g'(N^0)} \tag{3.2.10}$$

where τ appears as the effective relaxation time of the fluctuations. This equation has the simple solution

$$\langle \Delta N \rangle_{k-N^0} = (k - N^0)\exp\left(-\frac{u}{\tau}\right). \tag{3.2.11}$$

Inserting this solution into (3.2.8) we find the autocorrelation function of the fluctuations to be

$$\Delta\Phi(u) = \langle \Delta N(0) \Delta N(u) \rangle = \langle \Delta N^2 \rangle \exp\left(-\frac{u}{\tau}\right) \tag{3.2.12}$$

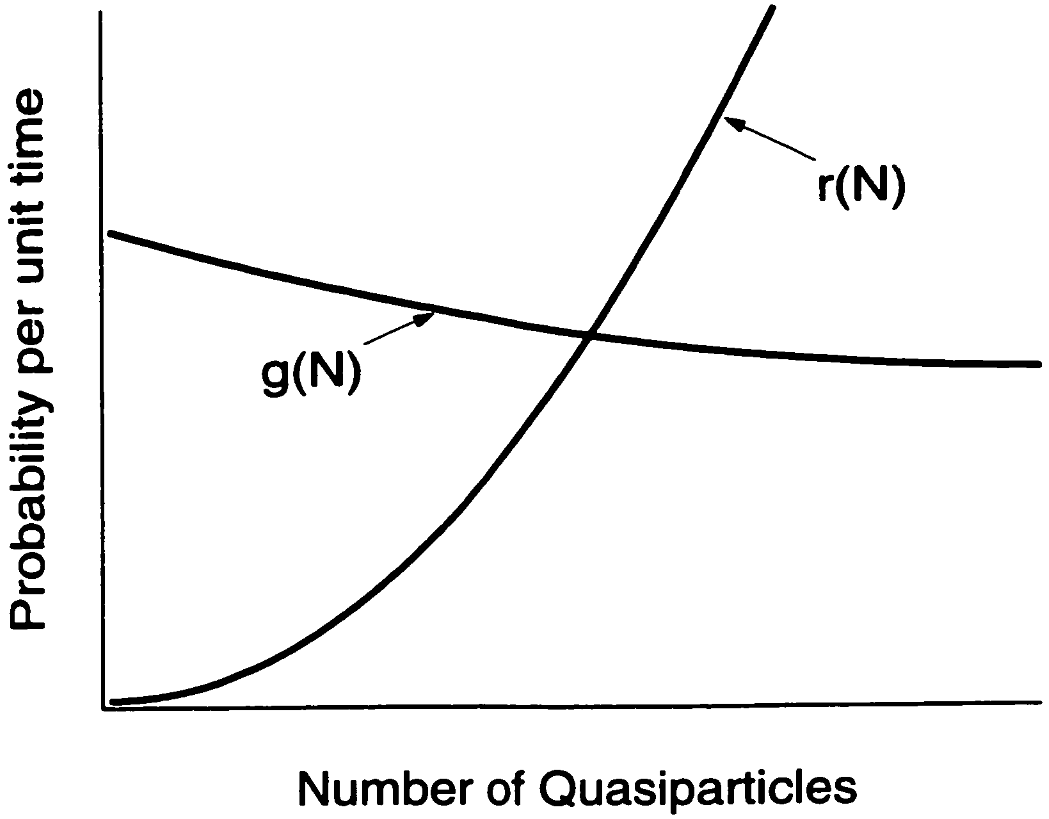


Figure 3.1 Sketch of the generation and recombination parameters, g(N) and r(N). The intersection of the curves yields the steady-state number of quasiparticles.

where $\langle \Delta N^2 \rangle$ is the variance of the fluctuations. We can then directly compute the power spectrum, G(ω), of the fluctuations as the Fourier transform of the autocorrelation function. We find

$$G(\omega) = \frac{4\langle \Delta N^2 \rangle \tau}{1+\omega^2\tau^2}. \tag{3.2.13}$$

We now have general expressions for the variance and power spectrum of the fluctuations in our two-level system. Before we specialize the equations more, we can make some general comments. First, if we combine (3.2.6) with (3.2.10), we find the much simpler expression for the variance of the fluctuations:

$$\langle \Delta N^2 \rangle = (\delta N)^2 \, r(N^0)\, \tau. \tag{3.2.14}$$

This says that the variance of the fluctuations is of order the number of particles that recombine in one correlation time. Now, looking at (3.2.10), we see that τ is inversely proportional to δN. This says the more quasiparticles that are lost (created) by a single recombination (generation) event, the faster the fluctuations. Also, looking at (3.2.10) we see that the time scale of fluctuations is inversely proportional the sum of the derivatives of the generation and recombination rates. This has a simple physical interpretation. In Fig. 3.1 we sketch the recombination parameter, r(N), and generation parameter, g(N), as a function of N. First, we note that the value of N where the curves intersect is the equilibrium value N^0. Next, we notice that for a stable system the derivative of r(N) will always be positive and the derivative of g(N) will always be negative. This is what maintains equilibrium. For example, if N fluctuates greater than N^0, then the recombination rate increases *and* the generation rate decreases. Both of these changes drive the system back to equilibrium. Even more, the steeper the change in the rates around equilibrium, the faster the system is driven back to equilibrium. This is why the time constants depend on the derivatives of r(N) and g(N) and why their contributions sum together.

To be able to apply the formulas derived above we must know what r(N) and g(N) are for our system. Luckily, if we already understand the dynamics of the system, it is general easy to deduce r(N) and g(N). In general, the rate equation of our system will be of the form

$$\frac{dN}{dt} = \delta N(g(N) - r(N)). \tag{3.2.15}$$

If we can derive or know an appropriate rate equation for our system, we can then read off g(N) and r(N).

We can consider, as an example, the case of simple generation and recombination of quasiparticles. By simple, we mean that quasiparticles are only lost to recombination with other quasiparticles and we ignore the effects of phonon trapping (which we will return to later). In this case, the two levels of our system are quasiparticles and Cooper pairs, with the total number of electrons constrained to be the normal state value. We will further assume that we are working at low temperatures and that the number of quasiparticles is small compared to the number of Cooper pairs. In general, we would expect g(N) to depend on the number of Cooper pairs. However, since the relative size of the fluctuations will be small compared to the number of Cooper pairs, we will assume g(N) is constant and equal to the equilibrium recombination rate. With that we can write the rate equation for our simple system as

$$\frac{dN}{dt} = 2\left(\Gamma_G - \frac{1}{2}\frac{R}{Vol}N^2\right) \tag{3.2.16}$$

where Γ_G is the constant generation rate, *Vol* is the volume of the system, and R is the recombination constant. The recombination constant is basically a constant of proportionality between the recombination rate and the number of ways to combine N quasiparticles, which is $N^2/2$.

From this rate equation, we can read off the parameters of our model:

$$g(N) = \Gamma_G \quad ; \quad r(N) = \frac{1}{2}\frac{R}{Vol}N^2 \quad ; \quad \delta N = 2. \tag{3.2.17}$$

We can then easily put these parameters into the equation above to find a familiar result for the variance of the fluctuations, $\langle \Delta N^2 \rangle = N^0$. We can also easily write down the power spectrum of the fluctuations

$$G(\omega) = \frac{4N^0\tau}{1+\omega^2\tau^2} \quad ; \quad \tau = \frac{Vol}{2RN^0}. \tag{3.2.18}$$

We see that the spectrum has a simple Lorentzian form with a bandwidth given by $1/\tau$.

3.3 Theory: Multivariable Master Equation

The simple one-variable master equation derived above is illustrative, but it is not sufficient to describe generation and recombination in a physical superconductor. In a thin-film superconductor, the phonon emitted when a pair of quasiparticles recombines can break another pair before the phonon escapes the film into the bath. This process, known as phonon trapping, extends the effective lifetime of a quasiparticle. We must therefore account for this process if we want to describe quasiparticle fluctuations in a physical superconductor. To

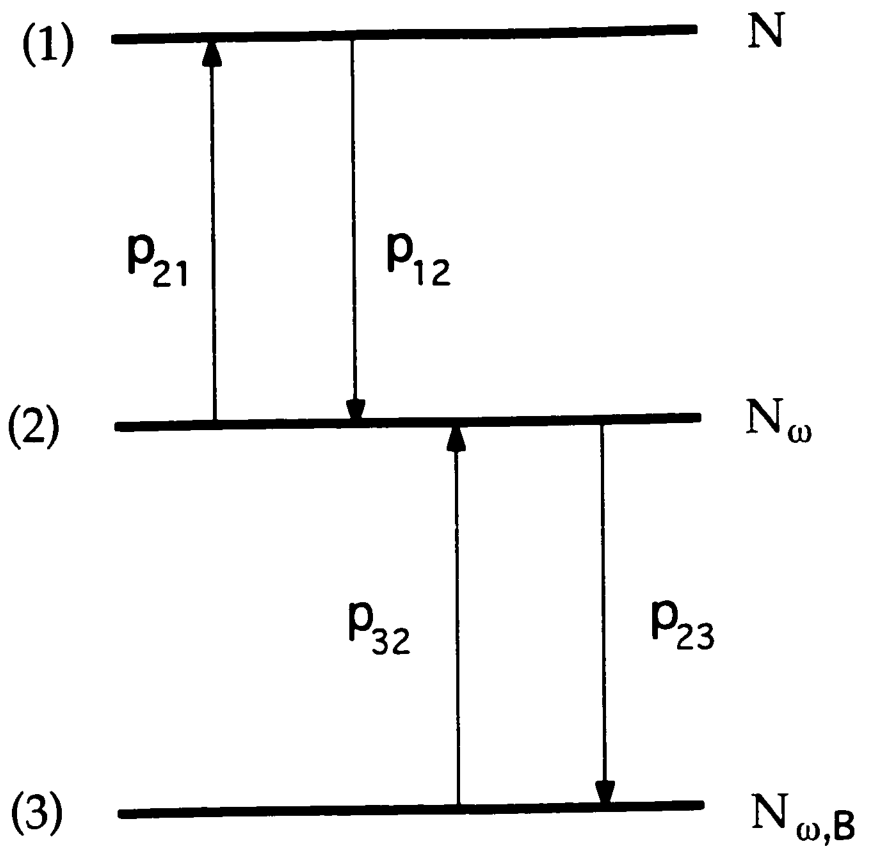


Figure 3.2 Schematic representation of our three level system. From top to bottom, the levels are quasiparticles, phonons in the electrodes, and phonons in the bath.

do this, we model our junction as a three level system, shown schematically in Fig. 3.2. The population in the three levels will be labeled by N, N_ω, and $N_{\omega,B}$ which are the number of quasiparticles in the electrodes, the number of phonons with energy $E_\omega > 2\Delta$ in the electrodes, and the number of phonons with $E_\omega > 2\Delta$ in the bath respectively. We only keep track of phonons with $E_\omega > 2\Delta$ because they are the only phonons that can generate new quasiparticles.

In the previous section, we thought of two quasiparticles recombining to form a Cooper pair, instead of quasiparticles recombining to form a phonon. In the end, however, N_ω is a more natural variable than the number of Cooper pairs for several reasons. From a statistical point of view, we can account for the recombination of two quasiparticles equally well as a transition to a Cooper pair or a transition to a phonon. From a dynamical point of view, however, keeping track of phonons is much more important then keeping track of Cooper pairs. As we will see shortly, the presence of phonons created by recombination can significantly change the effective recombination rate measured in experiments. On the other hand, the rate Γ_B at which phonons break pairs and generate quasiparticles *is* proportional to the number of Cooper pairs, but as long as the number of pairs is much greater than the number of quasiparticles, then Γ_B is approximately constant. Thus, we see that N_ω is a better choice.

We can describe the dynamics of the levels with the following system of three coupled differential equations:

$$\frac{dN}{dt} = 2\left\{-\frac{1}{2}\frac{RN^2}{Vol} + \Gamma_B N_\omega\right\} \tag{3.3.1}$$

$$\frac{dN_\omega}{dt} = \frac{1}{2}\frac{RN^2}{Vol} - \Gamma_B N_\omega - \Gamma_{ES} N_\omega + \Gamma_K N_{\omega,B} \tag{3.3.2}$$

$$\frac{dN_{\omega,B}}{dt} = \Gamma_{ES} N_\omega - \Gamma_K N_{\omega,B} \tag{3.3.3}$$

where Γ_{ES} is the rate at which phonons escape from the electrode to the bath and Γ_K is the rate at which phonons enter the electrode. The factor of 2 in the first equation comes from the fact that quasiparticles are generated and recombine in pairs. We have neglected the anharmonic decay of phonons as a loss process because it happens on a time scale much longer than phonon escape at these energies.

We can simplify these equations with the approximation that $N_{\omega,B}$ is constant, which is justified because the exchange of phonons with the junction is a very small perturbation to the bath. This simplification reduces (3.3.3) to the equality $\Gamma_B N_\omega^0 = \Gamma_K N_{\omega,B}^0$, where the superscripts indicate equilibrium values. We can then rewrite (3.3.2) as:

$$\frac{dN_\omega}{dt} = \frac{1}{2}\frac{RN^2}{Vol} - \Gamma_B N_\omega - \Gamma_{ES}\left(N_\omega - N_\omega^0\right) \tag{3.3.2'}$$

We then see that (3.3.1) and (3.3.2') are the well known Rothwarf-Taylor equations [Rothwarf 1967].

Following Gray [Gray 1981], we can linearize these equations for small perturbations by writing $N = N^0 + \Delta N$ and $N_\omega = N_\omega{}^0 + \Delta N_\omega$ and simplifying. If we let the vector $\mathbf{a} = (N, N_\omega)$ then we can write the linearized equations in matrix form

$$\frac{d(\Delta\mathbf{a})}{dt} = -\mathbf{\Gamma}\cdot\Delta\mathbf{a} \;\; ; \;\; \mathbf{\Gamma} = \begin{pmatrix} 2\Gamma_R & -2\Gamma_B \\ -\Gamma_R & \Gamma_\omega \end{pmatrix} \qquad (3.3.4)$$

where we have taken $\Gamma_\omega = \Gamma_B + \Gamma_{ES}$ and $\Gamma_R = RN^0/\mathrm{Vol}$ as the equilibrium recombination rate. The eigenvalues of $\mathbf{\Gamma}$ determine the time constants of the system's response to small perturbations. Gray showed that the dominant time constant for the quasiparticle response in the limit $\Gamma_R << \Gamma_B + \Gamma_{ES}$ is

$$\Gamma_R{}^* = 2\Gamma_R F_\omega^{-1} \;\; ; \;\; F_\omega = 1 + \frac{\Gamma_B}{\Gamma_{ES}} \qquad (3.3.5)$$

where F_ω is called the phonon trapping factor. It accounts for a phonon emitted by a recombination event breaking another pair before it escapes to the bath. We note that $F_\omega{}^{-1}$ is just the probability that a phonon escapes to the bath. $\Gamma_R{}^*$ is the time constant with which a small perturbation of the quasiparticle system will decay, and it is the rate we expect to measure in experiments. We see that the measured recombination rate, $\Gamma_R{}^*$, is in generally very different from the true equilibrium recombination rate Γ_R.

We can treat the fluctuations of this system with a multivariable master equation. The basic idea is the same as before, except we now describe the state of the system with levels (1) - (S) by a vector $\mathbf{a} = (N_1, N_2, \ldots, N_S)$ which represents the occupation of each level. In general only S-1 levels will be independent since the total number of excitations is constrained. We start by writing down the master equation for the process:

$$\frac{\partial P(\mathbf{a},t|\mathbf{a}',0)}{\partial t} = \sum_{\mathbf{a}''\neq\mathbf{a}} P(\mathbf{a}'',t|\mathbf{a}',0)Q(\mathbf{a};\mathbf{a}'') - \sum_{\mathbf{a}''\neq\mathbf{a}} P(\mathbf{a},t|\mathbf{a}',0)Q(\mathbf{a}'';\mathbf{a}) \qquad (3.3.6)$$

where $P(\mathbf{a},t\,|\,\mathbf{a}',0)$ is the probability that the system is in state **a** at time t given that it was in state **a'** at t=0, etc. and $Q(\mathbf{a},\mathbf{a}'')$ is the transition probability per unit time from state **a"** to **a**. Again, the first term says that the rate of change in the probability of finding the system in state **a** is the probability of it being one transition away from **a** times the rate of transition to **a**. The second term accounts for transitions out of state **a**. We can make this less abstract if we notice that the only allowed transitions in our system involve a single loss event in one level causing a creation event in a second level. We can then write

$$Q(\mathbf{a};\mathbf{a}'') = \left\{ \begin{array}{l} p_{ij} \;\; ; \; \mathbf{a}'' = \{n_1, \ldots, n_i, \ldots, n_j, \ldots\} \\ \qquad \mathbf{a} = \{n_1, \ldots, n_i - \delta n_i, \ldots, n_j + \delta n_j, \ldots\} \\ \\ 0 \;\; ; \; \text{otherwise} \end{array} \right\} \tag{3.3.7}$$

where δn_i is the "shot size" for each level.

We can then proceed along the same lines as the derivation in section **3.2**. I will not include the detailed derivation, instead quoting the results and referring to [van Vliet 1965] for a more detailed treatment. In analogy to the linearized time constant found in (3.2.10), we can write a linearized rate matrix, **M**, where the elements are:

$$M_{ij} = \delta n_i \sum_k \left(\frac{\partial p_{ik}}{\partial N_j} - \frac{\partial p_{ki}}{\partial N_j} \right) \Bigg|_{\{N_i\}=\{N_i^0\}} . \tag{3.3.8}$$

We can define a second matrix, **B** (which describes the second order Fokker-Plank moments), whose elements are:

$$\begin{aligned} B_{ii} &= \delta n_i^2 \sum_{k \neq i} (p_{ki} + p_{ik}) \approx 2 \delta n_i^2 \sum_{k \neq i} p_{ik}^0 \\ B_{ij} &= -\delta n_i \delta n_j (p_{ij} + p_{ji}) = -\delta n_i \delta n_j (p_{ij}^0 + p_{ji}^0) \end{aligned} . \tag{3.3.9}$$

We can then write the covariance matrix

$$\boldsymbol{\sigma}^2 = \langle \Delta \mathbf{a} \Delta \mathbf{a}^T \rangle = \tfrac{1}{2} \mathbf{M}^{-1} \mathbf{B} \tag{3.3.10}$$

where $\Delta \mathbf{a} = \mathbf{a} - \mathbf{a}^0$. We can also write the cross power spectrum matrix as

$$\mathbf{G}(\omega) = \frac{2}{\omega^2} \mathrm{Re} \left[\left(\mathbf{1} + \frac{\mathbf{M}}{i\omega} \right)^{-1} \mathbf{B} \right] \tag{3.3.11}$$

where Re[ ] means the real part and **1** is the identity matrix. The diagonal terms of **G** describe the power spectrum of the fluctuations of each level in the system.

The off-diagonal terms of **G** describe the cross power spectrum between the various levels. (The cross power spectrums are the Fourier transform of the correlation functions between the fluctuations of the various levels.) Each spectrum G_{ij} is a sum of individual Lorentzian spectra, like (3.2.18), with characteristic frequencies determined by the eigenvalues of **M**.

We can now specialize these results to our physical superconductor. Again, we will describe the superconductor as a three level system consisting of quasiparticles, phonons in the superconductor and phonons in the bath. The three levels are connected by various transitions labeled $\{p_{ij}\}$ in Fig. 3.2. Each transition represents a physical process that changes the occupation of the three levels. Transition p_{12} describes two quasiparticles recombining to create one phonon in the electrode. Transition p_{21} describes the reverse process, a phonon being absorbed and generating two quasiparticles. Transition p_{23} describes a phonon escaping from the electrode into the bath. Finally, p_{32} describes a phonon entering the electrode from the bath. We note that there is no direct connection between levels 1 and 3, the quasiparticles and the bath. Since we have a three level system, our underlying master equation is a two variable equation. We choose as our two variables the number of quasiparticles, N, and the number of phonons in the electrodes, N_ω. Referring to the rate equations for the system, (3.3.1) – (3.3.3), we can read off the transition probabilities, which we tabulate in Table 3.1.

Transition	**Symbol**	**Probability per unit time**
recombination	p_{12}	$(1/2)RN^2/Vol$
generation	p_{21}	$\Gamma_B N_\omega$
phonon escape	p_{23}	$\Gamma_{ES} N_\omega$
phonon entry	p_{32}	$\Gamma_{ES} N_\omega^0$

Table 3.1. Allowed transitions and the probability per unit time for each one.

In addition to the transition probabilities, we can also read off the shot size for each level, which is $\delta n_1 = 2$ for the quasiparticles and $\delta n_2 = 1$ for the phonons. Plugging these parameters into the above equations we find

$$\mathbf{M} = \begin{pmatrix} 2\Gamma_R & -2\Gamma_B \\ -\Gamma_R & \Gamma_\omega \end{pmatrix} \quad ; \quad \mathbf{B} = \Gamma_R N^0 \begin{pmatrix} 4 & -2 \\ -2 & 1+\frac{\Gamma_{ES}}{\Gamma_B} \end{pmatrix} \tag{3.3.12}$$

where $\Gamma_\omega = \Gamma_{ES} + \Gamma_B$. With these matrices we can then write the covariance matrix for our system. We find

$$\boldsymbol{\sigma}^2 = \begin{pmatrix} N^0 & 0 \\ 0 & N_\omega^0 \end{pmatrix} = \begin{pmatrix} N^0 & 0 \\ 0 & \frac{1}{2}\frac{\Gamma_R}{\Gamma_B} N^0 \end{pmatrix} \tag{3.3.13}$$

where we have used the principle of detailed balance to relate N_ω^0 to N^0. Thus, we again find that the variance of the occupation of each level is equal to the average occupation, as we expect from basic thermodynamic arguments. We also note that the off-diagonal terms are identically zero, implying that the quasiparticle and phonon fluctuations are independent. This is somewhat surprising since, as we will see later, the presence of the phonons does significantly modify the spectrum of the quasiparticle fluctuations.

Experimentally, we can only measure the spectrum of the quasiparticle fluctuations, so we will only calculate that spectrum. Using (3.3.11) and quite a bit of algebra, we obtain the quasiparticle spectrum

$$S(\omega) \equiv G_{11}(\omega) = \frac{2\alpha_1 \tau_1 N^0}{1+(\omega\tau_1)^2} + \frac{2\alpha_2 \tau_2 N^0}{1+(\omega\tau_2)^2} \tag{3.3.14}$$

where

$$\alpha_1 = 2\frac{\tau_1 - \tau_{ES}}{\tau_1 - \tau_2} \quad ; \quad \alpha_2 = 2\frac{\tau_{ES} - \tau_2}{\tau_1 - \tau_2}$$

and $\gamma_{1,2} = 1/\tau_{1,2}$ are the eigenvalues of **M** and $\tau_{ES} = 1/\Gamma_{ES}$. It is straightforward to show that if we integrate $S(\omega)$ over all ω we recover N^0 for the variance. This expression is completely general. However, in the limit $\Gamma_R << \Gamma_B + \Gamma_{ES}$ we can simplify the eigenvalues of **M** to $\tau_1 = 1/\Gamma_R^*$ and $\tau_2 = (\Gamma_{ES} + \Gamma_B)^{-1}$, where Γ_R^* is defined in (3.3.5). In this case, one time constant basically corresponds to the effective quasiparticle lifetime and one corresponds to the phonon lifetime. We can then interpret the first term of (3.3.14) as "intrinsicÓ quasiparticle fluctuations and the second term as phonon-driven fluctuations.

Now, in Al electrodes with our geometry and at the temperatures we are working we expect $\Gamma_R \sim 10^4\ s^{-1}$, $\Gamma_{ES} \sim 10^9\ s^{-1}$ and $\Gamma_B \sim 10^{10}\ s^{-1}$ [Kaplan 1976]. So, we are very much in the limit just described. In addition, we expect $\alpha_1 = 2(1 - 10^{-5})$ and $\alpha_2 = 2(10^{-5})$. This gives us a simplified expression for the spectrum:

$$S(\omega) \approx \frac{4\tau_R * N^0}{1+(\omega\tau_R *)^2}. \qquad (3.3.15)$$

If we compare this expression with the spectrum found in (3.2.19), we see that this power spectrum could have been obtained from a simpler one-variable master equation assuming effective generation and recombination parameters

$$r(N) = \frac{1}{2}\frac{R}{F_\omega Vol}N^2 \quad ; \quad g(N) = r(N^0) \qquad (3.3.16)$$

where the generation parameter g(N) is just a constant equal to the equilibrium recombination rate. This simplification is not general, but it is possible in our samples because the quasiparticle and phonon time scales are so widely separated. Basically, the quasiparticle system cannot respond to the fast phonon fluctuations and is only affected by the average number of phonons.

We note that comparing the two rate matrices in (3.3.4) and (3.3.12) $\mathbf{\Gamma} = \mathbf{M}$. As we have just seen, the eigenvalues of **M** determine the spectrum of the fluctuations. On the other hand, the eigenvalues of $\mathbf{\Gamma}$ determine the time constants of the dynamical response to small perturbations. The fact that these two matrices are equal in our model system imply that the timescales of the dynamical response and the fluctuations are the same. Before moving on to experimental results, we will take some time to explore the generality of this connection.

When we write rate equations like (3.3.1) – (3.3.3) we are making some implicit approximations. First, we approximate the occupation numbers of the levels, such as N and N_ω, as continuous variables, when they are in fact discrete variables. Second, we approximate the discrete and random transitions between levels as continuous and deterministic "flowsÓ. To understand the implications of these approximations, we start by deriving a differential equation for the expectation values of the level occupations from the master equation (3.3.6). We find the following system of equations:

$$\frac{\partial}{\partial t}\langle N_i\rangle = \sum_{j\neq i}\delta n_i\left(\langle p_{ji}(N_j)\rangle - \langle p_{ij}(N_i)\rangle\right) \tag{3.3.17}$$

where the indices i and j run over all levels. We can compare this equation to a general expression for the rate equations (similar to (3.2.15)), which is

$$\frac{\partial}{\partial t}N_i = \sum_{j}\delta n_i\left(p_{ji}(N_j) - p_{ij}(N_i)\right). \tag{3.3.18}$$

The only apparent difference is that we have dropped the expectation value brackets from the second system of equations. But, we must also keep in mind the subtle difference that the first equation is an exact differential equation for the continuous expectation value of a discrete variable. The second equation is only approximate for the reasons mentioned above.

However, in the special case where the $\{p_{ij}(N_i)\}$ are all linear functions of the occupation numbers,$\{N_i\}$, we then have that $\langle p_{ij}(N_i)\rangle = p_{ij}(\langle N_i\rangle)$ and we can actually interpret the rate equations as exact equations for the expectation values of the occupations of the levels. In many physical systems, although, the transition rates are at least quadratic in the occupation numbers, such that, $p_{ij} \sim N_i^2$ or $p_{ij} \sim N_iN_j$. In this case, we have, for example, that

$$\langle p_{ij}\rangle \sim \langle N_i^2\rangle = \langle N_i\rangle^2 + \langle \Delta N_i^2\rangle. \tag{3.3.19}$$

So, in the case of quadratic transition rates we can interpret the rate equations as approximate equations for the expectation values, ignoring terms of order the variance of the occupation number. In general though, we expect that $\langle \Delta N_i^2\rangle \sim \langle N_i\rangle$ and we can say that neglecting the variance terms is a valid approximation to order $\mathcal{O}(1/N)$. Thus, for a large system the rate equations actually describe the expectation values of the occupation numbers, to good approximation.

We can develop this idea a little further. If we take (3.3.17) and Taylor expand the transition probabilities to first order we get the following equation for small variations:

$$\frac{\partial}{\partial t}\langle\Delta N_i\rangle = \sum_{j\neq i}\delta n_i\left[\sum_k \frac{\partial p_{ji}}{\partial N_k}\langle\Delta N_k\rangle - \sum_k \frac{\partial p_{ij}}{\partial N_k}\langle\Delta N_k\rangle\right]_{\{N_k\}=\{N_k^0\}} = \sum_k M_{ik}\langle\Delta N_k\rangle \qquad (3.3.20)$$

where M_{ik} are the elements of the matrix **M** defined in (3.3.12). If we follow the same procedure for the rate equations, and we find that the linearized rate equations:

$$\frac{\partial}{\partial t}\Delta N_i = \sum_{j\neq i}\delta n_i\left[\sum_k \frac{\partial p_{ji}}{\partial N_k}\Delta N_k - \sum_k \frac{\partial p_{ij}}{\partial N_k}\Delta N_k\right]_{\{N_k\}=\{N_k^0\}} \equiv \sum_k \Gamma_{ik}\Delta N_k \qquad (3.3.21)$$

where have defined the linearized rate matrix **Γ**. We see, in general now, that **Γ** = **M** and that we can interpret the linearized rate equations as equations for the expectation values around equilibrium. This is the general connection between the fluctuations and the dynamics. The eigenvalues of **Γ** determine the dynamical response of the system to perturbations. The eigenvalues of **M**, in turn determine the characteristic times of the fluctuations. But, as we have just seen, the matrices are the same. Thus, the timescales measured from dynamic perturbations and from equilibrium fluctuations must be the same.

The proceeding discussion amounts to a statistical fluctuation-dissipation theorem for our system. In fact, we can derive the fluctuations of our system in a more conventional thermodynamic framework using the fluctuation-dissipation theorem [van Vliet 1965]. We start by considering the entropy function of our system, $S(E,Vol,\{N_i\})$, which is a function of the *extrinsic* thermodynamic variables, $\{a_i\}$, which are in this case the energy, E, the volume, Vol, and the particle occupation numbers, $\{N_i\}$. We can then define the *intrinsic* thermodynamic variables of the system as

$$X_k = \frac{\partial S}{\partial a_k}. \qquad (3.3.22)$$

In particular, for the extrinsic variables $\{N_k\}$, the corresponding intrinsic variables are $\{-\mu_k/T\}$, where μ_k is the chemical potential of k-th species of particles and T is the temperature. We can then approximate the probability of finding the system in a state **a** around the equilibrium state $\mathbf{a}^0$ as

$$P(\mathbf{a}) = C\exp\left[-\frac{1}{2k_B}\sum_{i,j} s_{ij}^0 \alpha_i \alpha_j\right] \tag{3.3.23}$$

where C is a normalization constant, $\alpha_i = a_i - a_i^0$ and we have introduced

$$s_{ij}^0 = \left.\frac{\partial^2 S}{\partial a_i \partial a_j}\right|_{\mathbf{a}=\mathbf{a}^0}. \tag{3.3.24}$$

The note that the matrix s_{ij}^0 arises from the Taylor expansion of $S(\mathbf{a})$ around the equilibrium state $\mathbf{a}^0$, where $S(\mathbf{a})$ takes on its maximum value. We also note that we can approximate the change of the intrinsic variables around equilibrium as

$$\Delta X_i = X_i(\mathbf{a}) - X_i^0 = -\sum_j s_{ij}^0 \alpha_j. \tag{3.3.25}$$

It is then straightforward to show that the variances of the fluctuations of the intrinsic variables are

$$\langle \alpha_i \alpha_k \rangle = k_B \left(s^0\right)_{ik}^{-1}. \tag{3.3.26}$$

We can understand this result in a straightforward fashion. The matrix $\mathbf{s}^0$ is the curvature of $S(\mathbf{a})$ around its equilibrium value. We then see that the magnitude of the fluctuations is inversely proportional to the curvature of $S(\mathbf{a})$.

Before we derive the frequency spectra of the fluctuations in this framework, we must introduce some concepts. If we consider $S(\mathbf{a})$ as the thermodynamic "potentialÓ for our system, in the sense that its maximum value and curvature determine the equilibrium behavior of our system, then we can consider its derivatives, the intrinsic variable $\{X_i\}$, as thermodynamic "forces.Ó We can then postulate a kinetic connection between our extrinsic variables and their corresponding forces. In steady state, the total thermodynamic force on the system must be balanced by dissipation in the system according to the relation

$$(X_i)_{total} = \sum_j R_{ij} \dot{\alpha}_j \tag{3.3.27}$$

where R_{ij} is the generalized resistance matrix and the dot indicates the derivative with respect to time. (The elements R_{ij} are also known as Onsager coefficients.) We can split the total force in (3.3.27) into an external "driving force" $V_i(t)$, acting on the system through the reservoir, and the internal restoring force, the intrinsic variables X_i. We can then rewrite (3.3.27) as

$$\mathbf{V}(t) = \mathbf{s}^0\boldsymbol{\alpha} + \mathbf{R}\dot{\boldsymbol{\alpha}} \cdot \tag{3.3.28}$$

where have changed matrix notation for convenience. If we take the Fourier transform of (3.3.28) we can characterize the response of the system to the external perturbation by the frequency dependent impedance matrix

$$\mathbf{Z}(\omega) = \mathbf{R} + \frac{\mathbf{s}^0}{i\omega} \tag{3.3.29}$$

where ω is the angular frequency. According to the fluctuation-dissipation theorem, the power spectrum matrix is then

$$\mathbf{G}(\omega) = \frac{4k_B}{\omega^2}\mathrm{Re}(\mathbf{Z}^{-1}) = \frac{4k_B}{\omega^2}\mathrm{Re}\left[\left(\mathbf{R} + \frac{\mathbf{s}^0}{i\omega}\right)^{-1}\right]. \tag{3.3.30}$$

This thermodynamic treatment is complete. The fluctuations can be calculated in this way, without appealing to the master equation formalism, if the entropy function can be derived and **R** can be deduced from (3.3.27).

We can, however, connect the thermodynamic result to the master equation approach. If we eliminate the external drive term from (3.3.28) we arrive at an equation for the intrinsic relaxation of the system:

$$\dot{\boldsymbol{\alpha}} = -\mathbf{R}^{-1}\mathbf{s}^0\boldsymbol{\alpha} \cdot \tag{3.3.31}$$

If we compare this to (3.3.20) we find the first connection

$$\mathbf{M} = \mathbf{R}^{-1}\mathbf{s}^0 \cdot \tag{3.3.32}$$

Comparing eqns. (3.3.10) and (3.3.26) for the variance of the fluctuations, we find

$$k_B\left(\mathbf{s}^0\right)^{-1} = \frac{1}{2}\mathbf{M}^{-1}\mathbf{B}. \tag{3.3.33}$$

Combining this result with (3.3.32) we find the final connection

$$\mathbf{R}^{-1} = \frac{1}{2k_B}\mathbf{B}. \tag{3.3.34}$$

Using (3.3.32) and (3.3.34) it is then a matter of simple algebra to show that the spectrum (3.3.30) is the same as the spectrum (3.3.11) derived using the master equation approach.

Thus, we see that behavior of our system can be derived either from a strictly thermodynamic point of view or through the master equation. We also see that the results of the two approaches are related in a simple fashion. This fact is important in and of itself, because it implies that important thermodynamic quantities can be deduced with only knowledge of the rate equations of a system, which may be easier to deduce than the entropy function, $S(\mathbf{a})$, or the Onsager coefficients, **R**.

3.4 Experiment

When we apply a bias voltage across our tunnel junction, the fluctuations of the number of thermal quasiparticles in the electrodes lead to excess current noise in the detector. At large bias, there is a simple connection between the number of quasiparticles in the box and the current. In Fig. 3.3, we show quasiparticles distributed in an energy range δE in the Al. (For a thermal distribution, δE is a few times k_BT.) Each quasiparticle is a coherent superposition of electron and hole, but it must tunnel as either a pure electron or pure hole. Biased at a voltage eV > δE, a quasiparticle can only tunnel from left to right as an electron, gaining energy eV. A quasiparticle cannot tunnel from left to right as a hole, because it would lose energy eV and tunnel into the gap on the right side. Similarly, quasiparticles can only backtunnel from right to left as a hole. Thus, for eV > δE, tunneling events from left to right and from right to left transfer a charge in the same direction and the associated currents add. The time-dependent current is then given by

$$I(t) = e\left(\frac{N_l(t)}{\tau_{tun}} + \frac{N_r(t)}{\tau_{tun}}\right) = e\frac{N(t)}{\tau_{tun}} \tag{3.4.1}$$

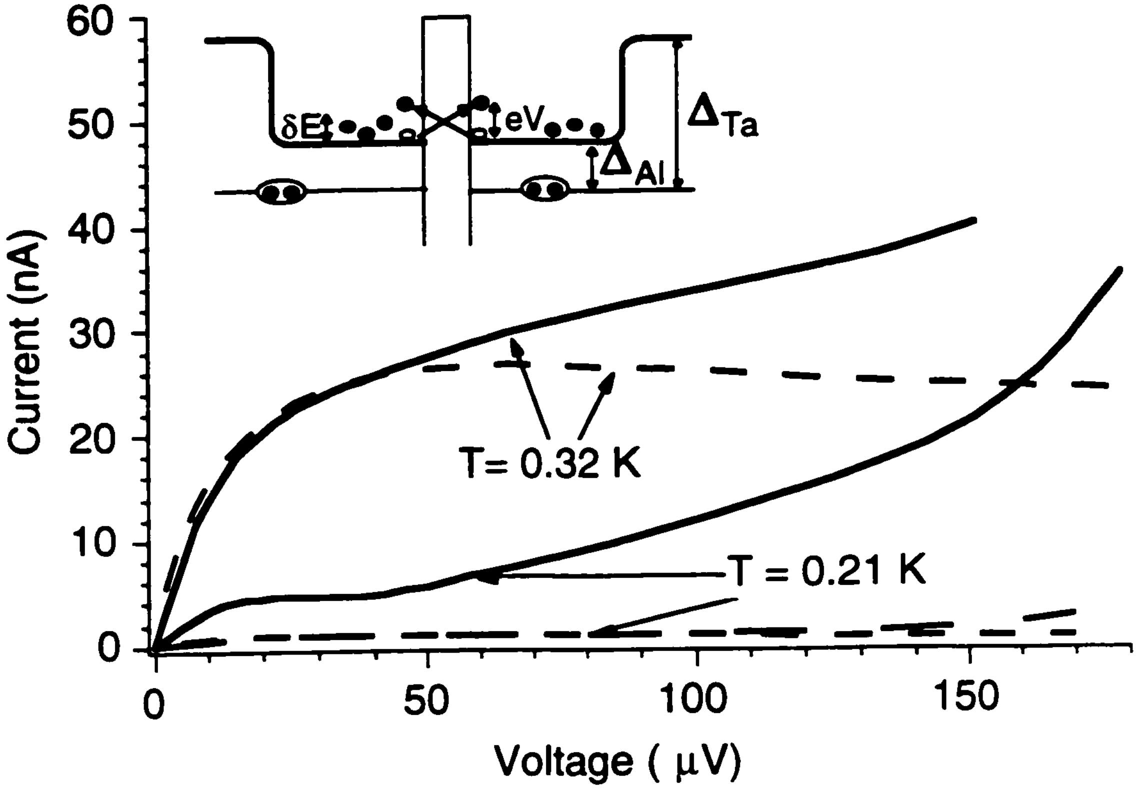


Figure 3.3 Current-voltage characteristics of the tunnel junction. The solid lines are I-V curves from device 1 that show the effect of heating. The long-dashed line is an I-V curve from a similar junction, device 2, without heating effects (see text). The short-dashed lines are BCS predictions. The low temp BCS curve and the curve from device 2 overlap. The inset is an energy band diagram of device 1 in the excitation representation.

where N_l and N_r are the numbers in the left and right side and τ_{tun} is the tunnel time [de Korte 1992]. The tunnel time here is twice the normal metal tunnel time because only half the tunneling processes are allowed. In writing (3.4.1) we have assumed that any variations in N(t) happen on a time scale $\tau >> \tau_{tun}$. As we will show later, the time scale of the fluctuations of the quasiparticles meet this condition.

At first, it might appear that the tunnel barrier divides the box into two distinct quasiparticle systems. However, it is a good approximation to treat the two halves as one box if the halves are strongly coupled. This condition for strong coupling is that $\tau_R^* >> \tau_{tun}$, where τ_R^* is the effective recombination time for the quasiparticles. If this condition is meet, a typical quasiparticle tunnels many times before it recombines and can interact with quasiparticles in both halves of the box [Gray 1978]. In our measurements, $n = \tau_R^*/\tau_{tun}$ is between ten and fifty, so that this strong coupling condition is always meet.

To measure the quasiparticle number with a tunnel barrier, we must use a voltage $eV > \delta E$. This introduces the problem of Joule heating. Each tunneling event dissipates eV of energy in the Al box, as seen in Fig. 3.3. In a tunnel junction with large Al leads the energetic quasiparticles diffuse away from the barrier and are replaced by "cool" quasiparticles from the leads. In our system, however, the energetic quasiparticles cannot diffuse away from the barrier because of the large-gap Ta leads. Thus, non-equilibrium heating occurs in the quasiparticle gas, including the generation of excess quasiparticles. Fig. 3.3 shows current-voltage (I-V) curves from two different devices along with theoretical BCS I-V curves. Device 2 (OPS-E00-7A) has a Ta lead only on the left side, so hot quasiparticles can diffuse away from the junction on the right side. Its I-V curve represents the behavior of a junction in equilibrium at the bath temperature and is quantitatively consistent with BCS predictions. The current of device 2 also scales with the junction area when compared to larger junctions we have made with one Ta lead. Device 1 (OPS-E00-6B), which has Ta leads on both sides and is used in the rest of the measurements in this section, shows excess current compared to device 2. The explanation for the excess current in device 1 is Joule heating, as described above (and in more detail in section **4.1**).

We directly measure the recombination time of quasiparticles in the box with single-photon absorption experiments [Wilson 2000]. A single photon from the mercury emission line at 4.89 eV (254 nm) is absorbed in the Ta on the left of Fig. 3.3, producing about 4000 quasiparticles. These quasiparticle diffuse to the Al where they can emit phonons and drop down in energy, becoming trapped. These trapped quasiparticles are a small perturbation to the $N_0 \sim 10^5$ steady-state quasiparticles in the Al box. The trapped quasiparticles circulate, tunneling and backtunneling, until they are lost to recombination, typically with a thermal quasiparticle. This produces a current pulse that decays exponentially on a time scale of the effective recombination time, τ_R^*.

Eqn. (3.4.1) implies that the fluctuations in the number of quasiparticles will cause fluctuations in the tunneling current with a spectral density

$$S_I(\nu)=\frac{4neI^0}{1+\left(\omega\tau_R^*\right)^2}=\frac{2(\alpha_1 n)eI^0}{1+\left(\omega\tau_R^*\right)^2} \qquad (3.4.2)$$

where I^0 is the steady-state current, $n = \tau_R^*/\tau_{tun}$ is the average number of times a quasiparticle tunnels before recombining, and the parameter α_1 is defined in (3.3.14). We predict $\alpha_1 = 2$ in Al at these temperatures. We have written the spectrum in this form to recall the magnitude of $S_I(\nu = 0)$ for Poisson shot noise of a tunneling current I^0, namely, $S_I(\nu = 0) = 2\, q_{eff}\, I^0$, where q_{eff} is the effective charge of the current carrier [Blanter 2000]. Equation (3.4.2) looks like a shot noise spectrum with an effective charge $q_{eff}/e = \alpha_1 n$. Our model predicts $q_{eff} \sim 100e$ at the lowest temperature. Poisson shot noise arises from the random timing of tunneling events. At high frequency we expect to recover the result $S_I(\nu >> 1/\tau_R^*) = 2e\, I^0$ in our junctions.

An additional source of noise could be a fluctuating charge imbalance between the hole-like and electron-like quasiparticle branches [Tinkham 1996]. In Al away from T_c, these fluctuations would be caused by quasiparticles randomly switching branches after scattering elastically from an impurity [Clarke 1986]. The branch mixing time in Al is of order $\tau_Q \sim 10^{-8}$ s [Clarke 1986], much faster then either the tunneling time, $\tau_{tun} \sim 10^{-6}$ s, or the effective recombination time $\tau_R^* \sim 10^{-4}$ s. This means that charge imbalance fluctuations should be averaged out in our measurements. (In addition, τ_Q should be independent of temperature in our operating range, compared to the exponential temperature dependence of τ_R^*.)

We have measured both the exponential falltime of single-photon pulses and the frequency spectrum of the current noise at different temperatures, as shown in Fig. 3.4. All measurements were made with the tunnel junction biased at V = 60 μV. The current pulse shown is the average of 500 single-photon current pulses. The decay is fit with a simple exponential to give one

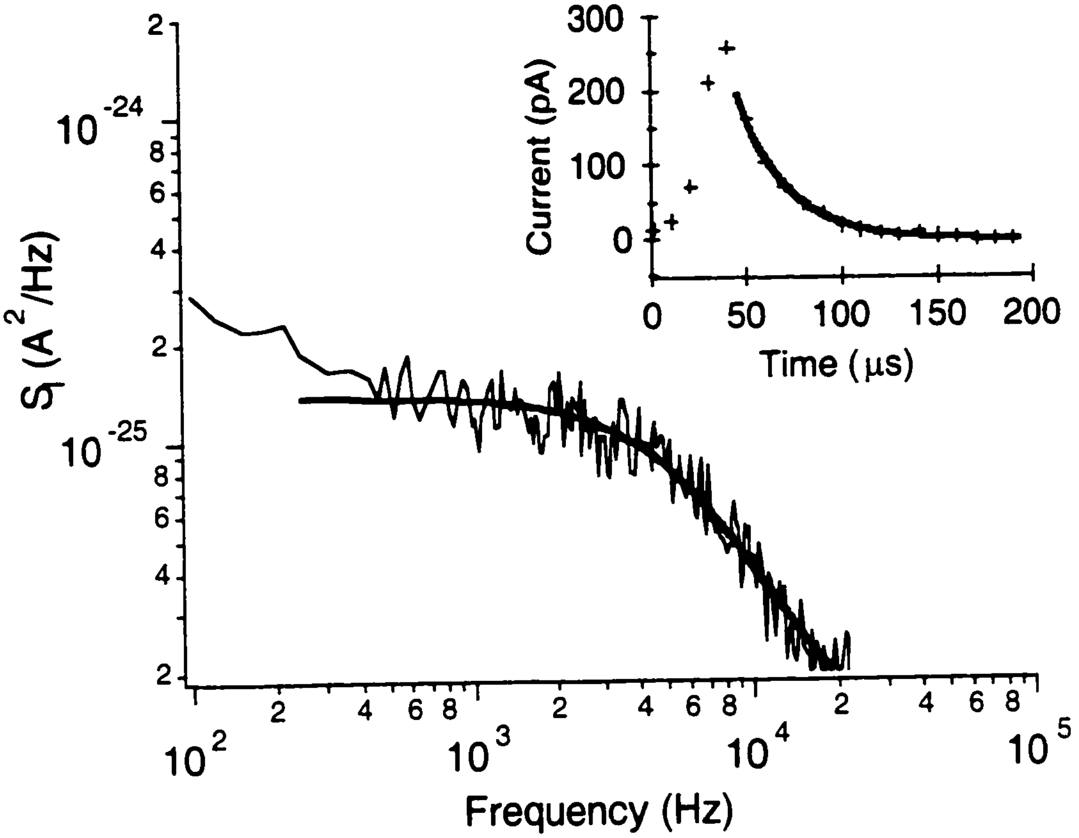


Figure 3.4 Spectral density, $S_I(\nu)$, of the tunneling current with a fit to the Lorentzian form (3.4.2). The inset is the average of 500 single-photon current pulses showing the decay with a fit to an exponential. For clarity, only a few percent of data points are displayed. T = 0.32 K.

measurement of τ_R*. The spectrum is the average of fast Fourier transforms of successive noise traces recorded with a digital oscilloscope. We fit a Lorentzian shape like (3.4.2) for a second measurement of τ_R*. Before fitting we digitally subtracted known electronic noise sources and removed lines caused by 60 Hz pick-up and microphonics. There is an additional white noise component that does not change with temperature and is about a factor of 15 less in power. We believe it is a combination of the conventional noise from pair and quasiparticle tunneling [Rogovin 1974]. The rise of the spectrum at very low frequency is due to the 1/f noise of the amplifier.

In Fig. 3.5 we plot both measurements of $\Gamma_R^* = 1/\tau_R^*$ versus the average number of quasiparticles in the box, $N^0(T)$. $N^0(T)$ is inferred from the average current using (3.4.1). Each N^0 (T) corresponds to a different bath temperature,

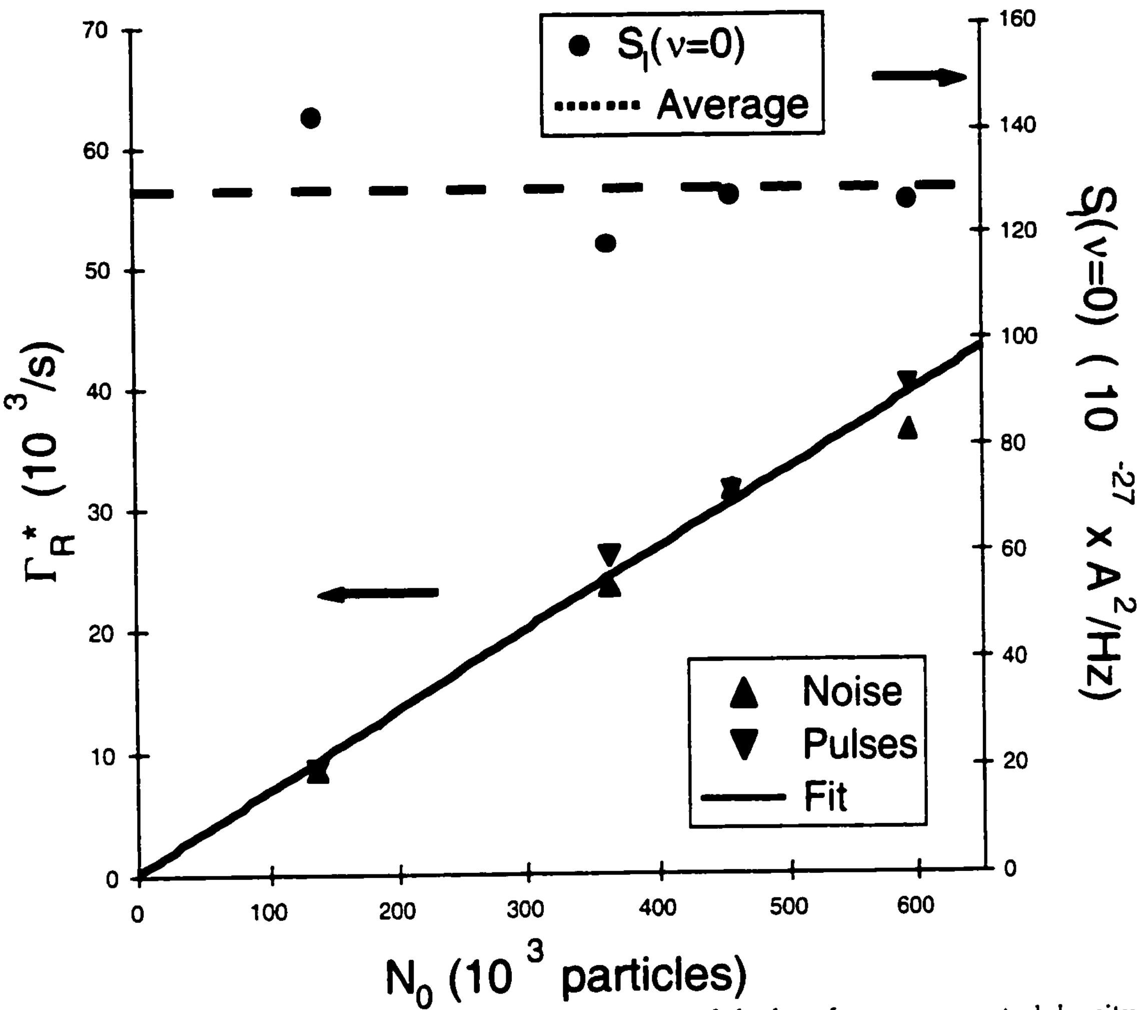


Figure 3.5 Data for the effective recombination rate, Γ_R^*, and the low-frequency spectral density, $S_I(\nu = 0)$. Γ_R^* was determined from the bandwidths of the spectra and the falltimes of the current pulses (Fig. 3.4).

between 0.21 K to 0.32 K. For the lower values of N^0 (T), the effective temperature of the quasiparticle gas was significantly higher than the bath temperature (refer to Fig. 3.3). The values of Γ_R^* inferred from the noise and the single-photon measurements agree, confirming the connection between the noise and the dynamics of quasiparticle loss. The solid line in Fig. 3.5 is a least squares fit to the total data set and is *not* constrained to zero intercept. We see that Γ_R^* varies linearly with N^0(T) and that the intercept of the line is very close to zero. These properties confirm that the loss rate Γ_R^* is in fact due to quasiparticle-quasiparticle recombination, as opposed to loss in traps, etc. [Gray 1971,1981]. The quality of the fit suggest that despite the fact that at the lowest bath temperature the quasiparticle gas is heated, it is behaving like it is in internal equilibrium at an effective temperature higher than the bath temperature. (The number of quasiparticles present at the bath temperature imply an effective temperature of $T_{eff} = 0.26$ K.)

The theory also predicts how the low frequency magnitude of the noise, $S_I(\nu = 0)$, should change as a function of temperature. Referring to (3.2.2), all factors in the magnitude of $S_I(\nu = 0)$ are approximately independent of temperature except I^0 and $\tau_R{}^*$. I^0 and $\tau_R{}^*$ both change with temperature because the number of quasiparticles changes. Specifically, $I^0 \sim N^0$ and $\tau_R{}^* \sim 1/N^0$. Thus, the product $I^0\tau_R{}^*$ is independent of N^0 (temperature) and therefore, $S_I(\nu = 0)$ should be independent of temperature. The right axis of Fig. 3.5 shows $S_I(\nu = 0)$, determined by fitting the spectra with (3.4.2), plotted versus N^0. We can see that there is *no* significant dependence of $S_I(\nu = 0)$ on N^0. Finally, the theory predicts the absolute magnitude of $S_I(\nu = 0)$, or equivalently the parameter α from (3.4.2). We measure a value $\alpha_I = 1.8 +/- 0.2$. This agrees well with our model prediction of $\alpha_I = 2$. The uncertainty quoted is one standard deviation of α_I and is dominated by the uncertainty in our knowledge of the electron density of states at the Fermi energy, $D(\varepsilon_F)$ [Friedrich 1997a].

An alternate explanation for the noise is that there is a flux of stray microwave or infrared photons incident on the device. The photons could come from the blackbody radiation of the dewar or from electromagnetic interference coupling down the leads to the device. If these stray photons have an energy $E_\gamma > 2\Delta$, then the photons will produce small single-photon pulses that are not individually resolved but still contribute noise [LeGrand 1997]. The frequency dependence of this noise would be the same as the fluctuation noise, so the data of Fig. 3.5 do not distinguish between these hypotheses. However, the change in the low frequency magnitude of the noise, $S_I(\nu = 0)$ as a function of temperature should be very different for the two noise sources. We expect the low frequency magnitude of noise from stray photons to be

$$S_{I,\gamma}(\nu=0)=2\beta\frac{\tau_R^*}{\tau_{tun}}eI_\gamma \qquad (3.4.3)$$

where I_γ is the photon induced current and β is a factor that accounts for the creation of quasiparticles in bunches by the photons. If the photons are monochromatic, then β would be the average number of quasiparticles created by a photon [LeGrand 1997]. Although this looks similar to (3.4.2), I_γ is independent of temperature. Thus, $S_{I,\gamma}$ should vary with temperature, being proportional to τ_R^*. The lack of such a dependence in Fig. 3.5 supports the interpretation that the noise is caused by thermodynamic fluctuations and not stray photons. In addition, the fact that our measurement of the parameter α agrees with the thermodynamic prediction of $\alpha_1=2$ suggest that stray photons do not cause the noise. If the noise arose from stray photons we would in general expect α_1 to be much larger (of order the average number of quasiparticles created by each photon). Finally, we have repeated our measurements of the noise in a different cryostat, in a light-tight copper sample holder only connected to room-temperature through SMA cables with cold, copper-powder rf-filters. The sample holder and the copper-powder rf-filters were all at the bath temperature. We observed no change in the noise under these conditions. These facts taken together conclusively rule out stray photons as the source of the noise.

3.5 Discussion

In principle, the fundamental time scale of electron-phonon interactions, known as τ_o, can be inferred from measurements of quasiparticle-quasiparticle recombination. The parameter τ_o is material dependent and is of general interest because the rate of many processes related to the interaction of electrons and phonons are determined by it [Kaplan 1976]. In particular, for a pair of quasiparticles at the gap edge, the expression for the recombination constant is

$$R=\left(\frac{2\Delta}{k_BT_c}\right)^3\frac{1}{2\Delta D(\varepsilon_F)\tau_o}. \qquad (3.5.1)$$

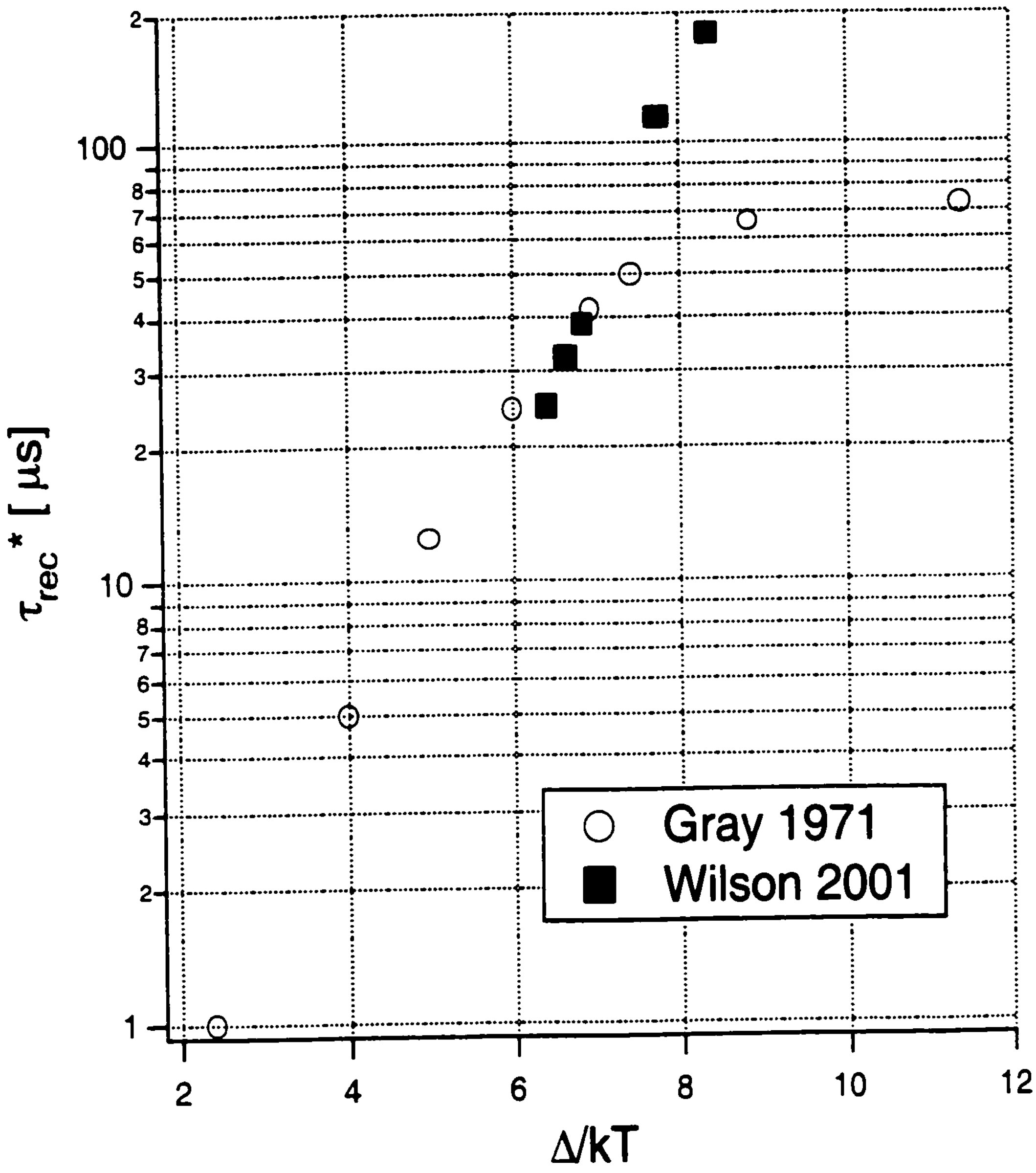


Figure 3.6 Comparison of quasiparticle lifetime measurements in this dissertation to previous measurements [Gray 1971]. Measurements by Gray were of Al on sapphire with Δ = 195 μV. Our films are on SiO_2 with Δ = 180 μV. Our data shows the lifetime following the theoretical dependence to lower temperature.

However, extracting τ_0 from recombination measurements at temperatures much less than T_c is complicated by phonon trapping. In fact, in the limit of strong phonon trapping, the measured recombination rate $\Gamma_R{}^*$ becomes

$$\Gamma_R^* = 2\frac{\Gamma_R}{\Gamma_B}\Gamma_{ES} \sim \Gamma_{ES} \tag{3.5.2}$$

because the pair-breaking rate, Γ_B, is also proportional to $1/\tau_0$. Thus, measurements of Γ_R^* in the presence of strong phonon trapping have no dependence on τ_0. If we were to ignore phonon trapping for the moment and insert our measured value of R into (3.5.1), we would extract a value for τ_0 of 1.65 μs. Numerous measurements of τ_0 in Al by other methods insensitive to phonon trapping find values of order 100 ns [Chi 1979], [Gray 1981]. This discrepancy suggests that our measurements are, in fact, in the limit of strong phonon trapping, so that they do not provide any independent information about τ_0. Our measurements do, however, confirm that the quasiparticle recombination rate is proportional to the quasiparticle density at lower temperatures and longer recombination times than previous experiments (refer to Fig. 3.6). Previous measurements showed recombination times that begin to deviate from the expected dependence at $T \sim 400$ mK and $\tau_R^* \sim 20$ μs and completely saturate at a maximum value of $\tau_R^* \sim 80$ μs below $T \sim 300$ mK [Gray 1971]. Quasiparticle loss into normal-metal regions created by trapped flux was proposed as the explanation for the deviation from theory in those measurements, although this explanation was not experimentally tested. Therefore, our measurements extend the range over which the basic physics of recombination has been verified in Al.

Current noise caused by fluctuations in the number of carriers has been observed previously in semiconductors. The term generation-recombination noise is commonly used to describe this noise. The treatment of the noise is quantitatively and qualitatively different for semiconductors. First of all, electrons in semiconductors recombine with holes, not other electrons. Quasiparticles in a superconductor combine with other quasiparticles. This difference significantly affects the statistics. In addition, the effect of non-equilibrium phonons, which is very important in thin-film superconducting devices, has not been treated in semiconductors.

From a technological point of view, generation-recombination noise may be more relevant to superconducting devices than to semiconductor devices. In intrinsic semiconductors, it is straightforward to cool devices to temperatures where there are no thermal excitations, so that the noise vanishes. Conversely, in extrinsic semiconductors, room temperature is sufficient to fully excite donors and acceptors, so that the noise is again suppressed. Neither of these condition is generally true in superconducting devices.

The effect of the fluctuations we have presented here imposes a previously unknown "thermodynamic limit" for superconducting tunnel junction (STJ) photon detectors. Previously, it was believed that only the statistical noise of the backtunneling process should limit the energy resolution for this type of STJ detector. The standard deviation of the backtunnel process, called the "backtunneling limit", is $(N_\gamma)^{1/2}$ referred to N_γ, the average number of quasiparticles created by an incident photon. On the other hand, the new thermodynamic limit we propose for this type of detector is $(\alpha_1 N^0/2)^{1/2}$ referred to N_γ. Basically, we see that the fluctuation of the thermal quasiparticles will degrade the energy resolution compared to the backtunneling limit if $N^0 \geq N_\gamma$. This is the case in our measurements, where $N^0 \approx 5N_\gamma$ in the best case. In our detectors, the backtunneling limit represents an energy resolution of about 0.2 eV referred to 5eV ultraviolet photons. On the other hand, the thermodynamic limit, using $\alpha_1 = 1.8$, is 0.45 eV independent of energy. We measure between 0.4-0.47 eV at from 2-5eV [Wilson 2001a]. This result is surprising. It had not been previously understood that backtunneling multiplication would also effectively multiply the shot noise due to thermally excited quasiparticles. It was believed that backtunneling could increase the signal of the detector over the noise of thermal quasiparticles to the point that the detector should be limited by the backtunneling noise of N_γ. This is clearly not the case. This generation-recombination noise could, in principle, be reduced by lowering the temperature of the junctions. However, this may not be straightforward since our lowest temperature is largely determined by self-heating and not the bath temperature.

Chapter 4: Physical Noise Processes: Master Equation Applications

4.1 The Master Equation

In Chapter 3, we derived expressions for the thermodynamic fluctuations of the quasiparticle number in our tunnel junctions using the master equation formalism. The master equation formalism itself is very general and can be applied to noise and fluctuations in a wide variety of physical systems. In this chapter we will apply the master equation to potential noise sources in our detectors: shot noise of a finite electrode and self-recombination. The standard derivation of shot noise in a tunnel junction assumes that the electrodes are ideal quasiparticle reservoirs. In our devices, however, the trap electrode is actually a small, isolated quasiparticle system. We expect that the random nature of the tunneling process itself will lead to fluctuations in the number of quasiparticles in the trap and, therefore, to excess current noise. Self-recombination is the process of two photon-excited quasiparticles recombining in the STJ. It is distinct from thermal recombination in which an excited quasiparticle recombines with a thermal quasiparticle. When the density of excited quasiparticles is high enough that a significant number of excited quasiparticles are lost to self-recombination, the response of the detector becomes nonlinear. In addition, self-recombination is a random process and should add noise to the detector. For both of these reasons, self-recombination imposes an upper limit on the useful energy range of an STJ.

As an introduction to the general application of the master equation, we start by considering a very simple system. We consider a one level population subjected to a linear loss process. A physical example of this would be thermal recombination of excited quasiparticles. In Chapter 3, we used a linearized equation to solve for fluctuations around an equilibrium value in our system. This example will illustrate that the master equation can be used to derive the complete time evolution of a probabilistic system. We start by writing the master equation for this system:

$$\frac{\partial P(N,t|N^0,0)}{\partial t} = -l(N)\cdot P(N,t|N^0,0) + l(N+1)\cdot P(N+1,t|N^0,0) \qquad (4.1.1)$$

where N is the number of quasiparticles and $l(N) = \Gamma N$ is the loss probability per unit time. The interpretation of this master equation is completely analogous to (3.2.1), except that there are no generation terms in this equation. If we multiply this equation by N and sum over all possible N, we get an equation for the expectation value of N:

$$\frac{d}{dt}\langle N\rangle_{N^0} = -\Gamma\langle N\rangle_{N^0} \qquad (4.1.2)$$

which has the straightforward solution

$$\langle N\rangle_{N^0} = N^0 \exp(-\Gamma t). \qquad (4.1.3)$$

This is the result we expect, namely, the number of quasiparticles decays exponentially with a characteristic time $1/\Gamma$.

As we discussed in Chapter 3, the result for the expectation value of N could have been derived by writing down a standard rate equation. What is unique about the master equation formalism is that it allows us to also derive the time evolution of the variance of N. Due to this, we multiply (4.1.1) by $\Delta N^2 = (N - \langle N\rangle)^2$ and sum over all values on N. We arrive at the following equation for the variance:

$$\frac{d}{dt}\langle \Delta N^2\rangle_{N^0} + 2\Gamma\langle \Delta N^2\rangle_{N^0} = \Gamma\langle N\rangle_{N^0}. \qquad (4.1.4)$$

This differential equation is slightly more complicated because it involves not only the variance, $\langle \Delta N^2\rangle_{N^0}$, but also the expectation value, $\langle N\rangle_{N^0}$. However, since we already have an explicit solution for the expectation value, we can solve (4.1.4) for the variance treating the expectation value as a "drive" term. If we assume that variance at t=0 is zero, meaning that the initial condition is known exactly, we get the following expression for the time evolution of the variance:

$$\langle \Delta N^2\rangle_{N^0} = N^0 \exp(-\Gamma t)(1-\exp(-\Gamma t)) = N^0 p_t(1-p_t) \qquad (4.1.5)$$

where we have defined $p_t = \exp(-\Gamma t)$. We can interpret this result in an interesting way. Basically, if we take a snap shot of the system at time t, the statistics of the system are described by an "instantaneous" binomial distribution with p_t as the instantaneous probability that a quasiparticle has been lost.

4.2 Self-Recombination

Self-recombination is the process of two photon-excited quasiparticles recombining to form a Cooper pair. It is one of several potential sources of quasiparticle loss in STJ detectors. It is similar to thermal recombination where an excited quasiparticle recombines with a thermal quasiparticle. In fact, the microscopic quasiparticle-quasiparticle interaction is identical in the two processes. We distinguish them because they have different effects on the response of the STJ. When the density of excited quasiparticles is much less than the density of thermal quasiparticles, the excited quasiparticles are much more likely to be lost to thermal recombination. Thus, the recombination probability is independent of the number of excited quasiparticles. This implies that thermal recombination is a linear loss process with respect to the detector response; it removes the same fraction of excited quasiparticles independent of incident photon energy. On the other hand, if the density of excited quasiparticles is much greater than the density of thermal quasiparticles, self-recombination is much more likely. In this case, the recombination probability *is* a function of the number of excited quasiparticles. Thus, self-recombination is a nonlinear loss process, meaning that a larger fraction of excited quasiparticles will be lost at higher incident energy.

Self-recombination may imposes a practical limit on the dynamic range of an STJ detector. Besides making the response of the STJ nonlinear, self-recombination is also a source of noise, due to its random nature. In this section, we will assess how stringent a limit self-recombination places on the practical energy range of an STJ. The noise associated with many loss and gain processes in an STJ can be calculated assuming that the process is binomial [Segall 2000]. This is not true of self-recombination. A binomial process is basically one where repeated, independent trials of an event have a fixed probability of success. For instance, thermal recombination of excited quasiparticles can be viewed as a binomial process since the probability of any quasiparticle being lost is independent of the fate of any other quasiparticle. On the other hand, for self-recombination the probability that a quasiparticle will be lost at a given time does depend on how many other quasiparticles have been lost.

We will consider self-recombination of quasiparticles in the trap electrode, ignoring thermal recombination. Quasiparticles can either be lost to self-recombination or they can tunnel to the counter-electrode. Quasiparticles that tunnel to the counter-electrode add to the collected charge and we assume for simplicity that they all diffuse away without backtunneling or reverse tunneling. We will keep track of the number of quasiparticles left in the trap, N_{trap}, and the number of quasiparticles that have tunneled and been collected, N_{coll}. We can write the following master equation for this system:

$$\begin{aligned}\frac{\partial}{\partial t}P(N_{trap},N_{coll},t) &= -\tau(N_{trap})\cdot P(N_{trap},N_{coll},t)+\tau(N_{trap}+1)\cdot P(N_{trap}+1,N_{coll}-1,t)\\ &\quad -r(N_{trap})\cdot P(N_{trap},N_{coll},t)+r(N_{trap}+2)\cdot P(N_{trap}+2,N_{coll},t)\end{aligned} \tag{4.2.1}$$

where $\tau(N)$ and $r(N)$ are the probabilities per unit time for tunneling and recombination respectively. In equation (4.2.1) we have dropped the notation indicating that the probabilities are conditional on a particular initial condition, but it is implied. From this master equation we can derive the following equations for the expectation values:

$$\begin{aligned}\frac{d}{dt}\langle N_{trap}\rangle &= -\frac{R^{*}}{V}\langle N_{trap}\rangle\left(\langle N_{trap}\rangle-1\right)-\Gamma_{tun}\langle N_{trap}\rangle\\ &= -\frac{R^{*}}{V}\langle N_{trap}\rangle^{2}-\Gamma_{tot}\langle N_{trap}\rangle\end{aligned} \tag{4.2.2}$$

$$\frac{d}{dt}\langle N_{coll}\rangle = -\Gamma_{tun}\langle N_{trap}\rangle \tag{4.2.3}$$

where we have used

$$r(N_{trap}) = \frac{1}{2}\frac{R^*}{V}N_{trap}\left(N_{trap}-1\right) \quad ; \quad \tau(N_{trap}) = \Gamma_{tun}N_{trap} \quad ; \quad \Gamma_{tot} = \Gamma_{tun} - \frac{R^*}{V} \qquad (4.2.4)$$

and V is the volume of the trap. Here we use the more complete expression for $r(N_{trap})$ compared to (3.2.17) because we are also interested in the behavior of the system when N_{trap} is small. As discussed in Chapter 3, we have neglected a term in (4.2.2) that depends on the variance of N_{trap}. The solution for the collected charge with the initial condition of N^0 quasiparticles in the trap is

$$\left\langle N_{coll}\right\rangle = N^0\frac{\Gamma_{tun}}{\Gamma_R}\ln\left(1+\frac{\Gamma_R}{\Gamma_{tot}}\right) \qquad (4.2.5)$$

where $\Gamma_R = R^*N^0/V$.

The master equation formalism also allows us to write differential equations for the variance of the occupation numbers. However, because $r(N_{trap})$ is quadratic in N_{trap}, the equations involve $\left\langle N_{trap}^3\right\rangle$. Since we do not have a priori knowledge about the relative size of $\left\langle N_{trap}^3\right\rangle$, we do not know if we can neglect terms involving it. If we do not neglect those terms, solving the equations for the variance entails also having to solve equations for $\left\langle N_{trap}^3\right\rangle$. However, the equations for $\left\langle N_{trap}^3\right\rangle$ involve fourth order moments, etc.. Thus, we cannot solve for the variance without having knowledge of the relative size of higher order moments of the occupation numbers.

We instead calculate the variance using a Monte Carlo simulation inspired by the master equation. The simulation involves explicitly calculating the time evolution of our model system. We start with N^0 quasiparticles in the trap and advance the simulation in small time steps. The time step is small enough that the probability of any event, either tunneling or recombination, is small. (In practice, the time step is constantly adjusted during the simulation to keep the probability between 0.1 and 0.01.) At each time step of length dt the probability of a tunneling event is calculated as

$$p_{tun} = \Gamma_{tun}N_{trap}dt \qquad (4.2.6)$$

where N_{trap} is the number of quasiparticles in the trap as a function of time. The probability of a recombination event is

$$p_{rec} = \frac{1}{2}\frac{R^*}{V} N_{trap}\left(N_{trap} - 1\right) \tag{4.2.7}$$

and the probability of nothing happening is

$$p_{nothing} = 1 - p_{rec} - p_{tun}. \tag{4.2.8}$$

At each time step a random number is drawn to decide what happens. A tunneling event decreases N_{trap} by 1 and increases N_{coll} by 1. A recombination event decreases N_{trap} by 2. This process is iterated until $N_{trap} = 0$. At the end of the simulation, we record the value of N_{coll}. We then repeat the simulation from the beginning with the same initial number, N^0. Each repeated trial ends with a different value for N_{coll}. By repeating the simulation many times, we can generate a distribution for N_{coll} and find its mean and variance. In addition, we can repeat the whole procedure for many different values of N^0 to find how the statistics of N_{coll} change.

We display the results from one set of simulations in Figures 4.1 – 4.4. The parameters that we have used in this simulation are $R^* = 3.36\ \mu m^3/s$, $V = 4.4\times10^{-3}\ \mu m^3$, and $1/\Gamma_{tun} = 12.5\ \mu s$. This value for R^* comes from our measurements in Chapter 3. The trap volume here is much smaller than any optical STJ would ever be. We have used such a small volume strictly to reduce the computational demands of the simulation. The computational time needed to complete the simulation scales with the initial number of quasiparticles, N^0. For a smaller junction, self-recombination becomes significant for smaller values of N^0. Thus, it takes less time to simulate smaller junctions. In addition, equation (4.2.5) implies that the functional form of the expectation values, at least, scale in a simple way with the size, suggesting that simulating a small junction should provide relevant information about larger junctions. This volume for the trap electrode would roughly correspond to a junction area of $0.04\ \mu m^2$, which implies

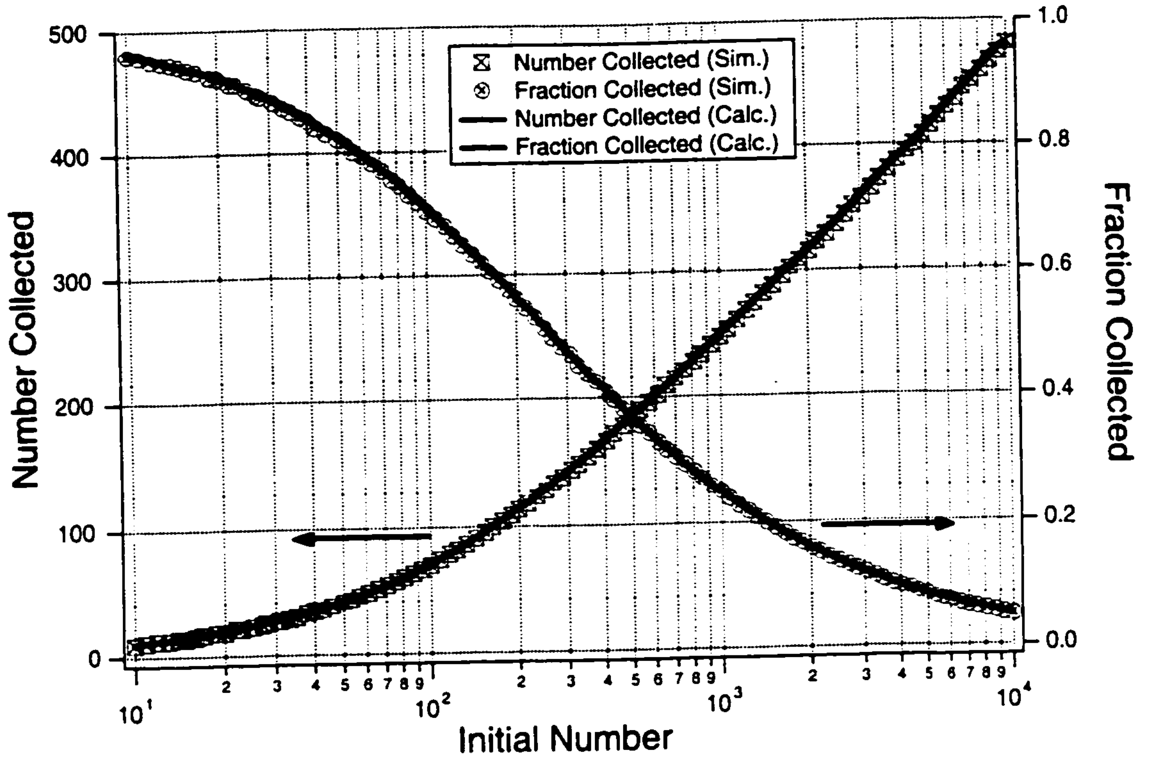


Figure 4.1 Results of Monte Carlo simulation of self-recombination process. On the left axis, the average value for the number collected is plotted versus the initial number in the trap. Plotted on the right axis is the average fraction collected, which is an estimate of the overall probability that a quasiparticle is collected. Results of the Monte Carlo simulation are plotted as markers while theoretical values are plotted as solid lines.

an optimal maximum energy of about 0.01 eV according to Chapter 5. A photon with an energy of 0.01 eV would create about 10 quasiparticles in the STJ.

In Fig. 4.1, we have plotted the mean N_{coll} versus N^0 from the simulation and for the master equation results (4.2.5). The agreement is very good. We can see that the nonlinearity introduced by self-recombination is not terribly strong. In fact, extending the incident energy (initial number) 1 order of magnitude above the "maximum" energy only reduces the collected fraction to about 70%. In addition, to reduce the collected fraction to less than 10% requires increasing the energy almost 3 orders of magnitude.

In Fig. 4.2, we have plotted the relative variance of N_{coll}, σ_{rel}^2, versus the fraction collected, which estimates the overall probability of being collected, p_{coll}. (The relative variance is the variance normalized to N^0.) Along with the simulation results, we have plotted the result for a simple binomial process,

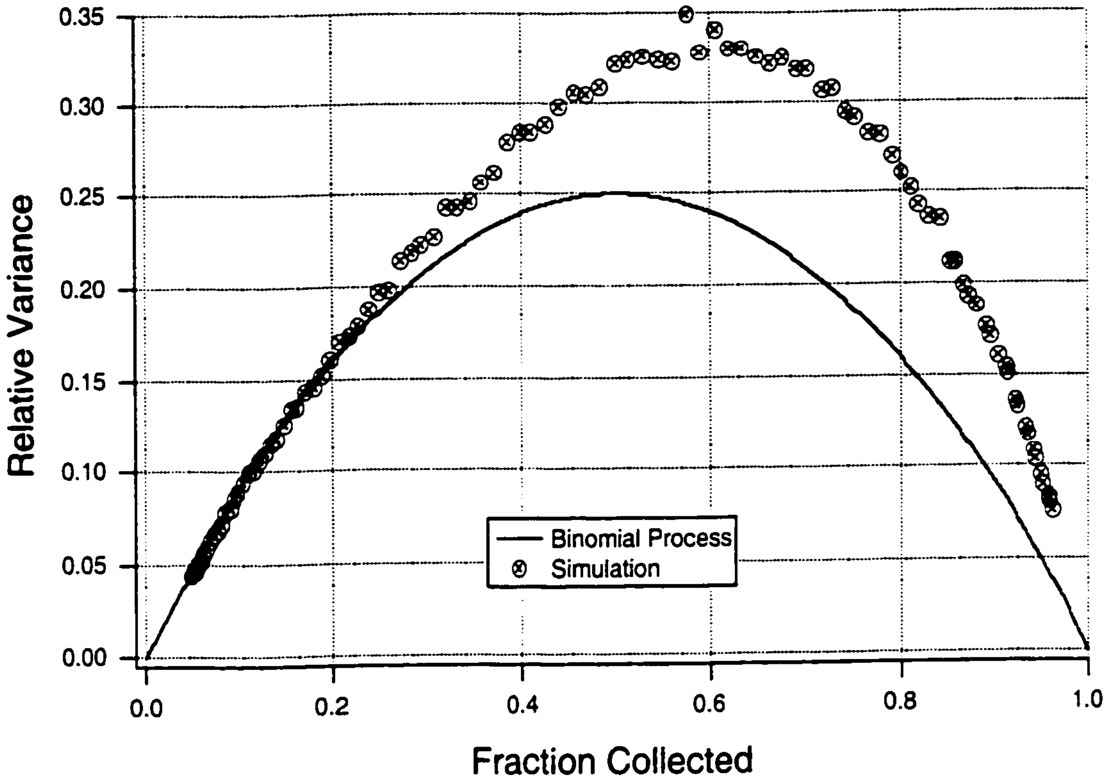


Figure 4.2 Relative variance of N_{coll} versus the fraction collected. The results of the simulation are compared to a simple binomial process.

namely, $\sigma_{bin}^2 = p_{coll}(1 - p_{coll})$. We see that this binomial result agrees when the fraction collected is small, but disagrees significantly when the fraction collected is large. In Fig. 4.3, we plot the ratio of the relative variance from the simulation to the binomial variance, $\sigma_{rel}^2/\sigma_{bin}^2$ versus the fraction collected. We have also plotted a quadratic fit to the ratio. What we see is that the ratio smoothly varies between about 1 for a small fraction collected and about 2 for a large fraction collected. We can understand these limits in the following way. When p_{coll}~1 the probability of self-recombination is small and each recombination event is approximately independent. Thus, we can approximate the process as a Poisson process with a shot size of 2. When p_{coll}~0, the probability of tunneling is small and we can approximate the process as a Poisson process with a shot size of 1.

To understand how important the noise due to self-recombination is, we can compare its contribution to the noise due to creation statistics, the Fano limit. Because self-recombination reduces the STJ charge output at the same time that it

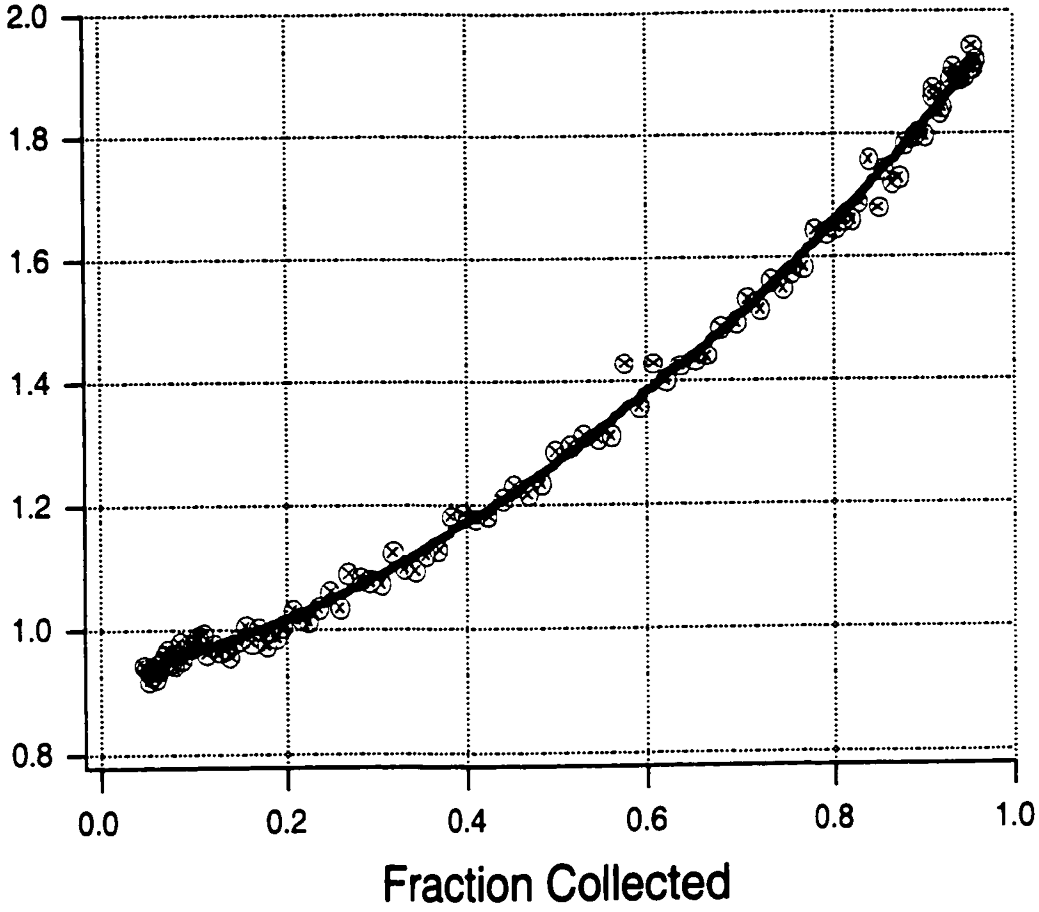


Figure 4.3 Ratio of the relative variance from the simulation to the relative variance of a simple binomial process. The result of the simulation is fit with a quadratic polynomial.

adds noise, we have to be careful when comparing the noise contributions. We choose to refer the noise to the "input" of the STJ, in other words, refer the fluctuations of the number to the initial number created [Segall 2000]. We can then write that the energy width due to self-recombination is

$$\Delta E_{self} = \sqrt{S\varepsilon E_{\gamma}} \quad ; \quad S = \frac{1}{p_{coll}}(1 - p_{coll})(0.92 + 0.30 p_{coll} + 0.76 p_{coll}^{2}) \qquad (4.2.9)$$

where E_{γ} is the energy of the incident photon, ε is the average energy to create one quasiparticle and we define S as the self-recombination factor analogous to the Fano factor, F. The coefficients of the quadratic term in (4.2.9) come from the fit in Fig. 4.3. We can define the dynamic range limit due to self-recombination as the energy where S=F, in other words, where the energy width increases by √2 due to self-recombination. If we assume F = 0.2, we find that the dynamic range limit corresponds to the energy where $p_{coll} = 0.9$. This implies that a nonlinearity of only 10% increases the noise by 40%. Thus, we conclude that the noise due to self-recombination is a much more stringent limit than that imposed by the nonlinearity itself.

4.3 Shot Noise of a Finite Electrode

The current across the tunnel barrier of the STJ flows in discrete "shots" as one quasiparticle after another tunnels. The time between each shot is random, so that even though there is a well-defined dc current, the current actually fluctuates if we look at short enough time scales. This phenomenon is called "shot noise." The standard derivation of shot noise in a tunnel junction assumes that the electrodes of the tunnel junction are ideal quasiparticle reservoirs, meaning that the tunneling does not perturb the occupation of energy levels in the electrodes [Rogovin 1974]. In such an ideal junction, the shot noise arises solely from the randomness in the timing of the shots. The trap of one of our STJs is certainly not an ideal reservoir. On the contrary, the trap is a finite (even small) quasiparticle system that is largely isolated by the Ta absorber. We expect that the number of thermal quasiparticles in the trap will fluctuate due to tunneling, possibly giving excess current noise. We note that this noise is related to the background of thermal quasiparticles. At sufficiently high bias voltage, tunneling of excited quasiparticles does not add charge noise [Segall 2001].

We can calculate the fluctuations of the quasiparticle number in the trap by simply applying the formulas of Chapter 3. We assume the junction is biased at a sufficiently high voltage such that quasiparticles can only tunnel as electrons from the trap and can only backtunnel as holes from the counter-electrode to the trap. For now, we will also assume that the counter electrode is an ideal reservoir such that the probability per unit time of a backtunneling event is constant. To apply the formulas of Chapter 3, we map this tunneling problem to a 1 variable generation-recombination problem. The one occupation number is N_{trap}. With respect to N_{trap}, we see that tunneling from the trap is equivalent to recombination and backtunneling to the trap is equivalent to generation. We can therefore write the "recombination" probability per unit time, $r(N_{trap})$, and "generation" probability per unit time as

$$r(N_{trap}) = \Gamma_{tun} N_{trap} \quad ; \quad g(N_{trap}) = \Gamma_{tun} N_{trap}^{0} \tag{4.3.1}$$

where $g(N_{trap})$ is just a constant equal to the average value of $r(N_{trap})$. We can then apply equation (3.2.6) to calculate that the variance of the fluctuations is $\langle \Delta N_{trap}^{2} \rangle = N_{trap}^{0}$. In addition, can apply (3.2.13) to calculate that the power spectrum of the fluctuations is

$$G(\omega) = \frac{4 N_{trap}^{0} \tau_{tun}}{1 + (\omega \tau_{tun})^{2}} \tag{4.3.2}$$

where $\tau_{tun} = 1/\Gamma_{tun}$.

We now want to calculate the excess current noise generated by these number fluctuations. In Chapter 3, we calculated the current noise by assuming that the tunneling current instantaneously followed the fluctuations of the number. We made that assumption in that case because the time scale of the number fluctuations was much longer than the tunnel time. That is clearly not true here. Still, if we follow the same procedure here we find the following excess current noise spectrum:

$$S_{I}(\omega) = \frac{2eI^{0}}{1 + (\omega \tau_{tun})^{2}}. \tag{4.3.3}$$

While this result is attractive in its simplicity, the preceding discussion does not properly motivate it.

We can, however, give a rigorous proof of the form of the noise spectrum in the following way. We start by considering a tunnel junction that has two finite electrodes which we still call the trap and counter electrode. We now consider the tunneling current produced by adding a *single* quasiparticle in our system. The quasiparticle will tunnel back and forth between the two electrodes producing a delta function of current each time it does. After backtunneling from the counter electrode to the trap, the quasiparticle will remain in the trap for a characteristic time $\tau_{tun} = 1/\Gamma_{tun}$ before tunneling again. Similarly, after tunneling to the counter electrode, the quasiparticle will remain there for a characteristic time $\tau_{CE} = \tau_{tun}(V_{CE}/V_{trap})$. Thus, the tunneling current is described by two intermixed Poisson processes. We can calculate the noise spectrum of such a process for arbitrary values of the ratio V_{CE}/V_{trap} [Devoret]. However, we are only interested in the limit $V_{CE}/V_{trap} >> 1$, because it is in this limit that the counter electrode behaves as an ideal reservoir. In this limit, we find that the noise spectrum of this current is

$$\begin{aligned} S_1(\omega) &= 4e^2 \frac{\Gamma_{tun}\Gamma_{CE}}{\Gamma_{tun}+\Gamma_{CE}}\left[1+\frac{1}{1+(\omega\tau_{tun})^2}\right] \\ &= 4e^2\Gamma_{tun}\frac{V_{trap}}{V_{trap}+V_{CE}}\left[1+\frac{1}{1+(\omega\tau_{tun})^2}\right]. \end{aligned} \tag{4.3.4}$$

Now we imagine that the electrodes are filled with a uniform density of quasiparticles, such that the total number in the system is $N_{tot} = n^0(V_{trap} + V_{CE})$, where n^0 is the density. If we assume every quasiparticle is independent, then each one will produce a tunneling current identical to what we have just described. We can then simply add the noise contributions of all the quasiparticles to find the noise spectrum of the filled electrode, $S_I(\omega) = N_{tot}S_1(\omega)$. We find:

$$\begin{aligned} S_I(\omega) &= 4e^2 n^0 V_{trap}\Gamma_{tun}\left[1+\frac{1}{1+(\omega\tau_{tun})^2}\right] \\ &= 2eI^0\left[1+\frac{1}{1+(\omega\tau_{tun})^2}\right] \end{aligned} \tag{4.3.5}$$

remembering that $I^0 = 2en^0V_{trap}\Gamma_{tun}$ because the average current from the counter electrode must balance the average current from the trap. At high frequency, this spectrum is that of a standard shot noise process. However, we see that there is also excess noise at lower frequencies with exactly the form (4.3.3).

To help confirm this result, we can also calculate the current noise spectrum using a Monte Carlo simulation. We can easily simulate the fluctuations of N_{trap} in an analogous way to the previous section. In addition, we simulate the current noise in the following way. First of all, we keep the time step of the simulation, dt, constant. At each time step we calculate the time-dependent probability of a tunneling event as $p_{tun} = \Gamma_{tun} N_{trap} dt$. The probability of a backtunneling event is $p_{back} = \Gamma_{tun} N_{trap}{}^{0} dt$ and is constant. The probability that nothing happens is $p_{nothing} = 1 - p_{back} - p_{tun}$. Again a random number is drawn at each time step to decide what happens. In this case, we keep track of two time series. The first is $N_{trap}(t)$ and the second is I(t). A tunneling event decreases N_{trap} by 1 and adds a delta function of current (with an integrated area of e-) to I(t). A backtunneling event increases N_{trap} by 1 and also adds a delta function of current to I(t). If nothing happens in the time step, N_{trap} does not change and I(t) is set to zero. We run the simulation for a certain number of iterations and at the end calculate the power spectrum of the number fluctuations and the current noise using the fast Fourier transform (FFT). We can also repeat the simulation many times, producing different time series each time, and average the resulting power spectra.

As a test of the simulation's ability to properly simulate the current noise spectrum, we can run it holding both the tunneling and backtunneling probabilities constant. With this constraint, the simulation should produce a white shot noise spectrum. We have plotted the output of this test in Fig. 4.5. The parameters of this simulation were $N_{trap}{}^{0} = 500$, $\tau_{tun} = 2.5\ \mu s$, $dt = 250$ ps, iterations=$2^{20} \sim 10^{6}$ and 1000 averages. With these parameters the probability of tunneling and backtunneling events are $p_{tun}=p_{back}=0.05$. We see that the simulation does indeed produce a white noise spectrum. However, the magnitude is somewhat too small. The expected value is $S_I=2eI^0=2.05\times10^{-29}$. We instead find that the magnitude is $S_I=1.85\times10^{-29}$. We see that the magnitude is low by 10% and note that this percentage is similar to the total probability of a shot in each time step, $p_{shot} = p_{tun}+p_{back}$. Thus, we could make an empirical guess that the

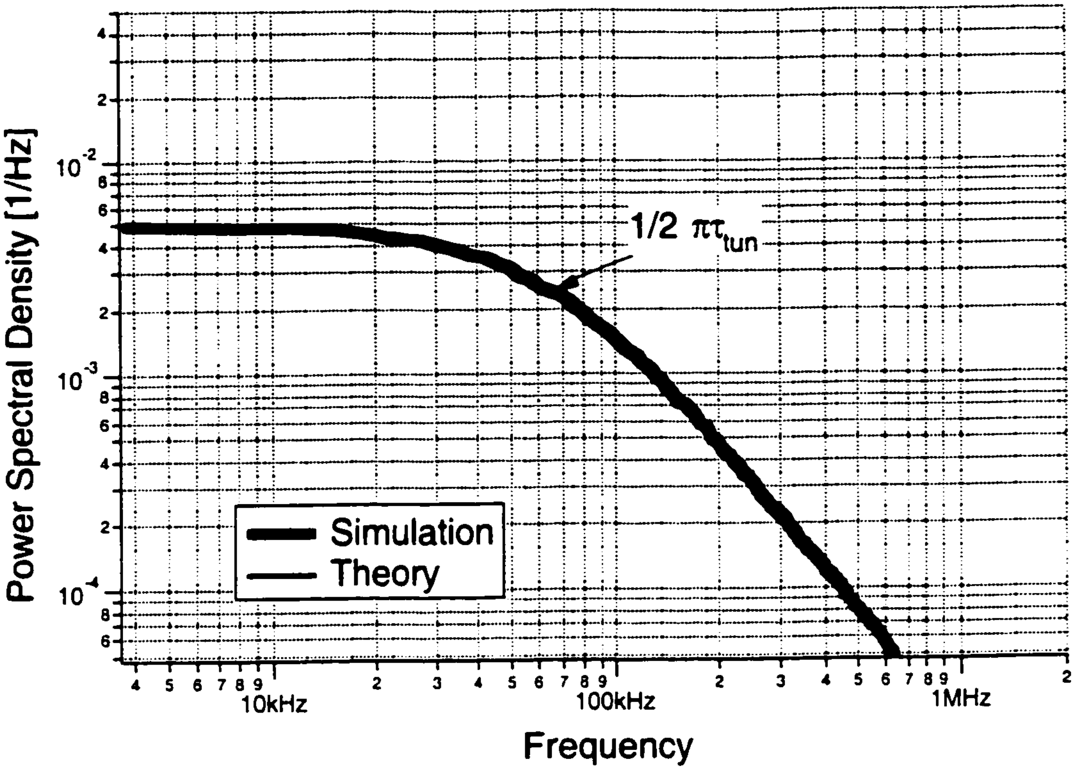


Figure 4.4 Simulated and calculated power spectrum of the fluctuations of the number of quasiparticles in the trap. The average number of quasiparticles in this simulation was 500.

error in the shot noise power of the simulation is first order in p_{shot}. This is reasonable because if $p_{shot} = 0.1$ it means that dt is one tenth of the average time between shots. Thus, the "granularity" of our time axis is about 10% and the timing of the shots is only random at a 10% level. We could decrease dt to decrease the granularity of the time axis. However, we would also need to increase the number of iterations in the time series (and therefore computational time) to keep the total time, T = iterations*dt, constant. The total time of the simulation sets the frequency resolution of the spectrum, i.e., df = 1/T. The frequency resolution also determines the minimum observable frequency and we need to have df << Γ_{tun} if we want the simulation to reveal any interesting features.

With some understanding of its computational limitations, we can move on to the full simulation. We have run the simulation with the same parameters as above but now allowing N_{trap} and p_{tun} to vary in time. (We also increased the

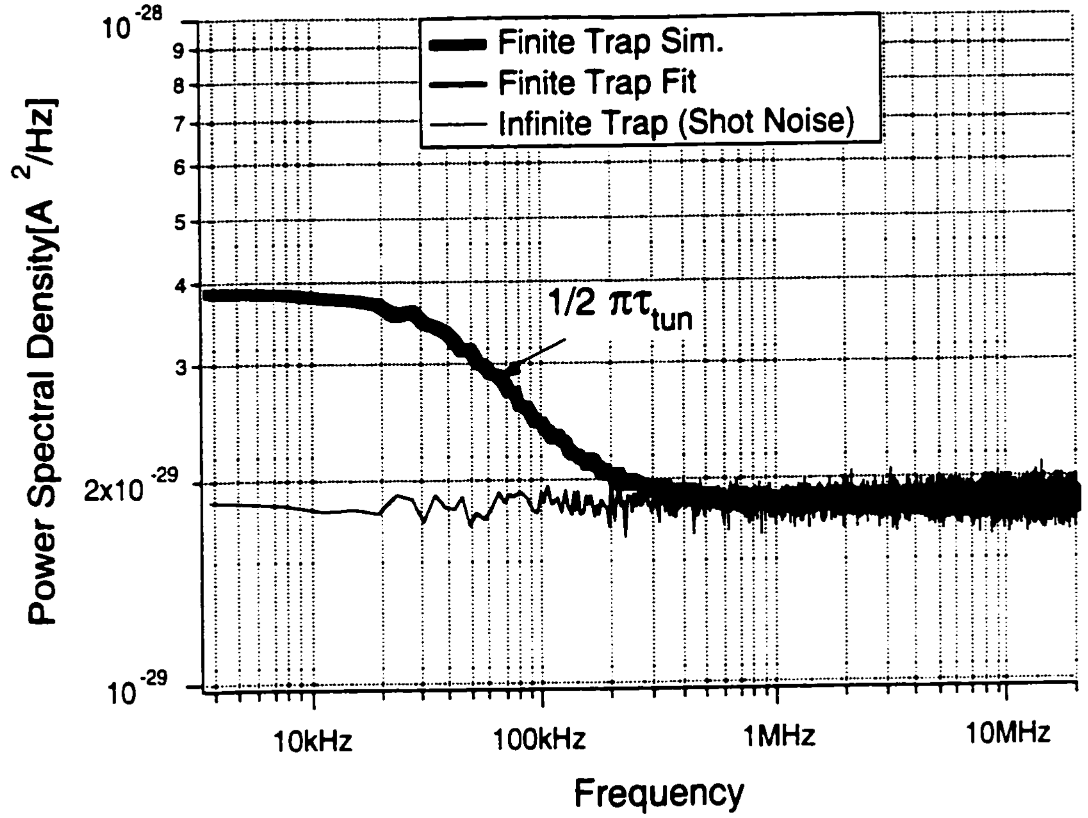


Figure 4.5 Power spectrum of current noise simulations. We have plotted the spectrum generated when N_{trap} is held constant and when it is allowed to fluctuate. The average number of quasiparticles is 500 in both cases.

number of averages to 10,000.) In Fig. 4.4 we have plotted the resulting power spectrum of the number fluctuations of N_{trap} along with the exact theoretical curve (4.3.2). We see that they agree well in both magnitude and shape. If we integrate this curve we should obtain the variance of the fluctuations $N_{trap}{}^0$=500. We find in fact 490, which is about a 2% error.

In Fig. 4.5 we plot the power spectrum of the current fluctuations. We see that the magnitude of the noise at high frequency agrees well with the output of the standard shot noise simulation, matching the 10% error. We also see that there is definitely excess noise below a characteristic frequency. To extract how much excess noise there is and what its characteristic frequency is, we fit the spectrum with the following curve:

$$S_{Fit}(\omega;\tau,H,L) = L \bullet \frac{eI^0}{1+(\omega\tau)^2} + H \bullet eI^0 \tag{4.3.6}$$

where the fitting parameters are τ, L, and H. The fit is displayed in Fig 4.5 and the parameters extracted are $\tau = 2.5\ \mu s = \tau_{tun}$, L = 1.95, and H = 1.8. The difference in H from the expected value of 2 agrees with the 10% we observed for the shot noise simulation. We see that simulated excess noise also agrees well with the predicted spectrum (4.3.5).

The form of the calculated spectrum (4.3.5) is independent of N_{trap}^{0}. To see if simulation reproduces this behavior, we can lower the average number of quasiparticles in the simulation. In particular, we have run the simulation with the parameters $N_{trap}^{0} = 1$, $\tau_{tun} = 2.5\ \mu s$, dt = 12.5 ns, iterations = $2^{15} \sim 32{,}000$ and 1000 averages. We note that here $p_{tun}^{0} = p_{back} = 0.005$ despite the fact the dt is much larger and the number of iterations is much smaller. This is true because N_{trap}^{0} is so much smaller. We plot the current noise spectrum for this simulation in Fig. 4.6 along with a fit to (4.3.4). The parameters for the fit $\tau = 2.5\ \mu s = \tau_{tun}$, L = 1.98 +/- 0.02 , and H = 1.975 +/- 0.0005, where the errors quoted are the uncertainties in the fit. We see that the magnitude of the high frequency part of the spectrum is off by 1%, which is what we expect because $p_{shot} = 0.01$. We also see that the shape of the excess noise agrees very well with (4.3.5). In fact, this time L = 2 within the error of the fit. Thus, the shape of the simulated spectrum is not sensitive to the number of quasiparticles.

We conclude that the current noise of a tunnel junction with one finite electrode is described by (4.3.5). The result is that for frequencies less than $1/2\pi\tau_{tun}$, the current noise power is doubled with respect to the noise from two ideal electrodes.

We have not made an in-depth study of the current noise in our non-backtunneling STJs. We do however observe that the current noise power, S_I, from these STJs is about 4 times what we expect for standard shot noise. It is possible that part of the excess noise is due to number fluctuations in the trap. If, in addition, there is a tunnel barrier between the counter-electrode and the wiring (see discussion in Chapter 5) number fluctuations in the counter-electrode

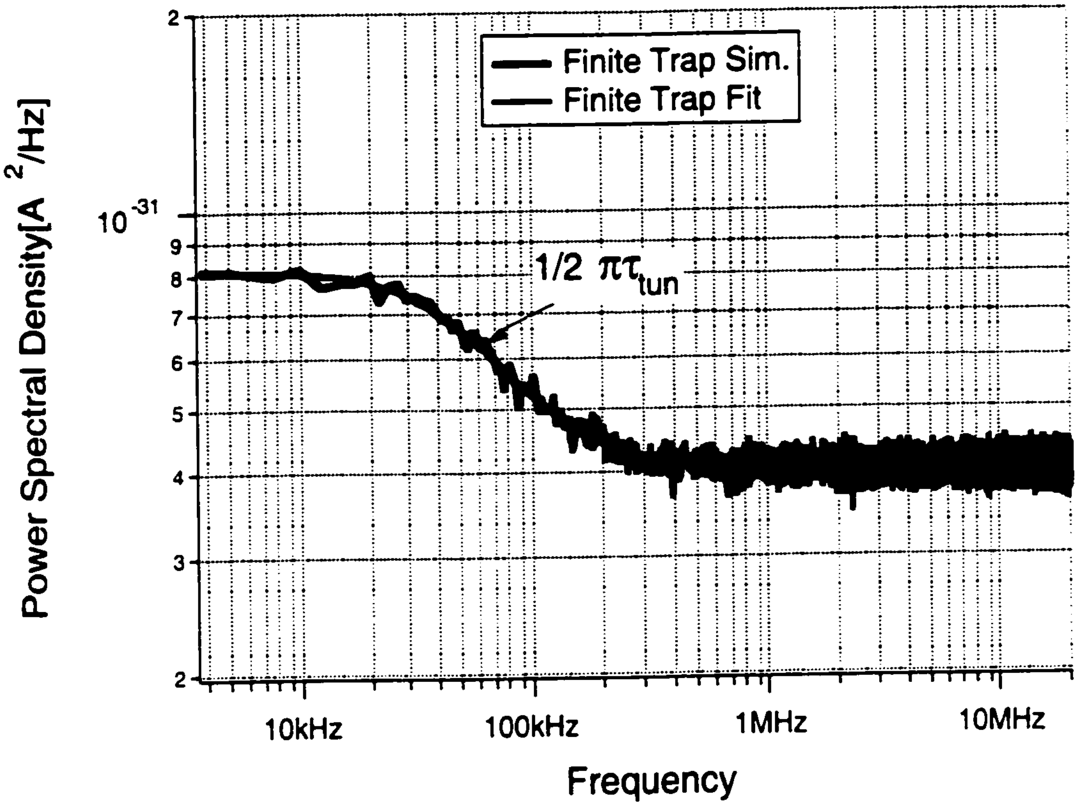


Figure 4.6 Power spectrum of current noise simulation. The average number of quasiparticles in the trap is 1.

could account for more noise. Detailed measurements of the noise spectrum would have to be made to answer the question definitively.

Chapter 5: Device Performance

5.1 Device Types

We present measurements of several different types of devices. The most important distinction is between backtunneling devices and non-backtunneling devices. Backtunneling devices have a Ta plug interrupting the Al wiring of the tunnel junction (refer to Fig. 1.1). The Ta plug prevents the diffusion of excited quasiparticles away from the junctions, promoting backtunneling and charge gain. Non-backtunneling devices omit the Ta plug, allowing excited quasiparticles to freely diffuse away from the junction, minimizing backtunneling. We have characterized both types of devices in the optical/UV energy range. Another important distinction is between a device with imaging capability and a single-pixel device. Imaging devices have two tunnel junctions contacting opposite sides of a strip absorber and can resolve many spatial pixels along the strip. Single-pixel devices have one tunnel junction contacting the absorber.

The optimum size for the tunnel junction of an STJ detector is determined by the energy range for which the STJ is intended. In general, the electronic noise of the detector plus amplifier will be minimized by making the tunnel junction smaller: a smaller junction has higher dynamic resistance, smaller capacitance and less shot noise. On the other hand, if a junction is too small, excited quasiparticles will recombine with each other (a process called self-recombination) leading to a nonlinear response and excess noise. Thus, the optimum junction should be as small as possible without leading to significant self-recombination at the maximum energy. In the limit of weak self-recombination, the fraction lost is

$$f_{lost} = \frac{1}{4}\frac{\Gamma_R^{\bullet}}{\Gamma_{tun}} = \frac{1}{4}\left(\frac{2R^{\bullet}N_0}{Vol}\right)\left(2e^2 D(\varepsilon_F)\cdot Vol\cdot R_{NN}\right) = 0.02\frac{E_\gamma}{A_J}$$

where N_0 is the initial number of quasiparticles in the junction, R_{NN} is the normal-state resistance of the junction, E_γ is the photon energy in eV and A_J is the area of the junction in μm^2. (In computing the numerical factor of 0.02 we have assumed our standard critical current density of 30 A/cm^2, a trapping multiplication of 2 and $R^* = 3.36\ \mu m^3/s$.) We note that the fraction lost to recombination is independent of the electrode volume because both $\Gamma_R{}^*$ and Γ_{tun} are inversely proportional to volume. We want the fraction lost to be less than $f_{lost} = 0.1$ so that the noise due to self-recombination does not dominate other sources [Section **4.2**]. This implies that the optimum tunnel junction size is 0.2 μm^2/eV. Thus, the best tunnel junction size for a detector designed to cover the optical/UV spectrum (up to about 5 eV) is 1 μm^2. Unfortunately, producing a junction this small is well outside the capabilities of our standard photolithography process. In fact, the smallest junction area that we have been able to produce reliably is 100 μm^2. (The limiting step in the lithography process is lifting off the via in the SiO insulating layer.) Thus, all of the measurements presented in this chapter were done with 100 μm^2 junctions. The shape of the junctions is a diamond with one axis 20 μm long and the other 10 μm long. Based on our lithography tolerances, the smallest junction that can be made with this process is ~ 35 μm^2. We have made junctions this small with low yield, although we believe the yield could be improved in the future.

5.2 Backtunneling Devices

The first devices that we measured were backtunneling devices. We expected that the charge gain associated with backtunneling would help us overcome the tremendous technical challenge of reading out the small charge , $Q_\gamma < 1$ fC, created by an optical/UV photon. Our designs did lead to large charge gain in the devices. However, these devices also displayed excess current noise and self-heating, as detailed in Chapter 3. For these reasons, these devices did not ultimately provide good energy resolution. Still, measuring and

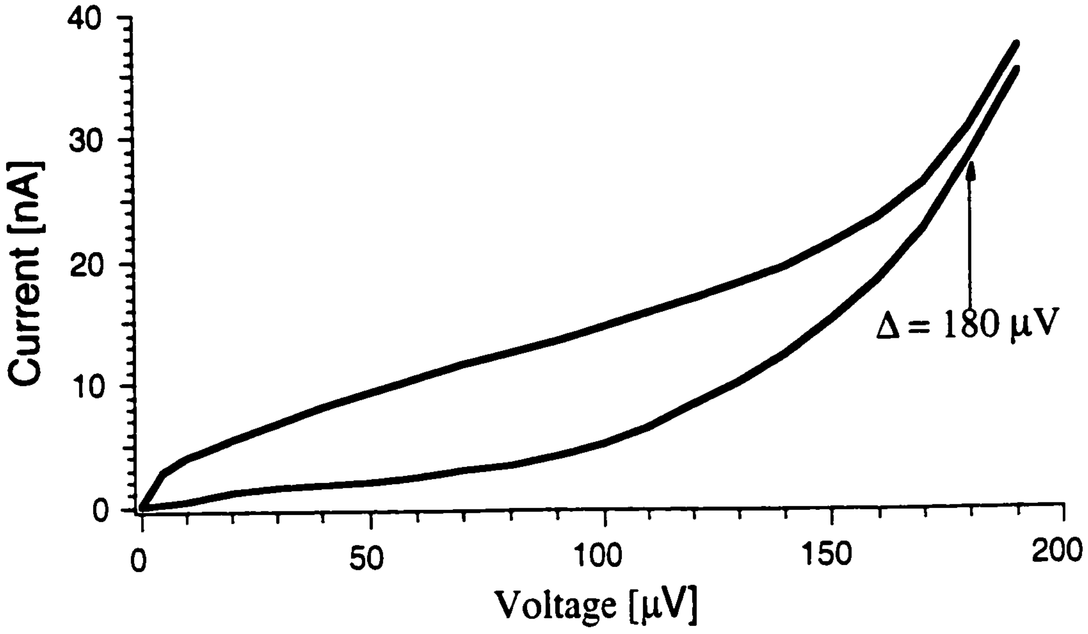


Figure 5.1 I-V curves of backtunneling device OPS-F99-6A after different laser warming cycles.

understanding these devices allowed us to refine our measurement setup and techniques, laying a foundation for subsequent progress.

Figure 5.1 shows a set of I-V curves for a 100 μm^2 junction that is part of a backtunneling STJ, device OPS-F99-6B. These measurements were all made at a base temperature of 0.21 K. The most important observation is that the subgap current is much higher and the dynamic resistance is much lower than what is expected from BCS predictions. For the top curve, the subgap current at 100 μV is about 12 nA, compared with a BCS prediction of about 1 nA. The dynamic resistance is about 8 kΩ, compared with the prediction of greater than 1 MΩ. We have made many junctions of this size and larger sizes that did show near BCS behavior. There are many things, including junction defects and trapped flux, that can lead to excess current in a tunnel junction. We are confident, however, that the behavior of these devices is caused by self-heating of the junction. This self-heating is a systematic property of our backtunneling geometry. At sufficiently high bias voltage, V, a quasiparticles gains an energy eV each time it tunnels. In a tunnel junction with no Ta plug, these hot quasiparticles can diffuse away from the junction. However, in our backtunneling devices, these quasiparticles are confined on both sides by high gap Ta. Thus, all of the bias energy is dissipated in the relatively small volume of the Al electrodes. This causes the junction to heat above the bath temperature, creating excess quasiparticles.

There are a number of empirical facts that suggest self-heating as the source of the excess current. First, the junctions of all functioning backtunneling devices that we have fabricated have shown this excess current phenomenon. In contrast, junctions of non-backtunneling devices (without Ta plugs) fabricated on the same wafers, all have I-V curves very close to the BCS prediction. There is also evidence that the excess current is caused by the presence of excess quasiparticles in the junction, as opposed to leakage or other things. As we stated in the previous chapter, the recombination rate of excited quasiparticles is linearly dependent on the number of quasiparticles. We have measured the recombination rate of photon-excited quasiparticles at different bias voltages and found that it depends linearly on the bias current. The recombination time ranges from $\tau_{rec} = 180\ \mu s$ for $I_{bias} = 4.3$ nA to $\tau_{rec} = 80\ \mu s$ for $I_{bias} = 11.8$ nA. (See chapter 3 for a description of the measurement technique.) This implies that the increase in current with increasing bias voltage is due to an increase in the number of steady-state quasiparticles. The final piece of evidence is that we have reprocessed some backtunneling devices, which had all shown the excess current, and fabricated a second Al wiring layer on them that allowed the quasiparticles to bypass the Ta plugs. The converted devices then all showed currents and resistances close to BCS predictions.

At a microscopic level, we believe that the increase in the number of quasiparticles is directly related to multiple tunneling. Most often, a quasiparticle will tunnel only once before it scatters inelastically and emits a phonon with energy $E \sim eV < \Delta$. Even if this process is repeated many times, it will not lead to an increase in the number of quasiparticles because none of the emitted phonons will be able to break a pair. These phonons with energy, $E < 2\Delta$, pass out of the Al film and away into the substrate. However, occasionally a quasiparticle may tunnel multiple times before it scatters inelastically. If it tunnels enough times before scattering to have an energy above the gap greater than 2Δ, it can then emit a phonon with sufficient energy to break a pair. We do not have a detailed simulation of this multiplication process, but preliminary calculations show that this explanation is plausible [Kozorezov 2002].

We found that we could affect the magnitude of the excess current in some devices. In the course of a single cryogenic run, we would warm the junctions above their transition temperatures by coupling light from a laser pointer to the devices. We found that if we did this several times during a single run we would often measure different amounts of excess current. Figure 5.1 contains sample I-V curves of device OPS-F99-6B, recorded after repeated cycles of laser heating in one cryogenic run. The amount of excess current evident in the I-V curve did not change if we did not warm the junctions. We observed that, in most cases, the recombination time measured by photon excitation increased if the excess current decreased, increasing the charge gain due to backtunneling. This suggests that the number of excess quasiparticles was reduced. (However, in the two cycles that we found the lowest currents, we found that the measured recombination time and charge gain were low.) We also observed that the amount of excess current *noise* decreased with decreasing current. For those reasons, we generally repeated the laser cycling until we found the device in a low current state with more charge gain and less noise. Clearly, this rather haphazard way of preparing the junctions for operation is undesirable.

We do not understand the mechanism for the variation of excess current. A possibility is that laser cycling the junctions produced varying configurations of fluxons trapped in the Al electrodes (though not in the junction itself). The normal metal cores of the fluxons can serve as trapping centers for quasiparticles. Trapping into a normal metal may be an efficient way to remove quasiparticles that self-heating adds. Recombination is a poor way to removes these quasiparticles because the recombining pair emits a phonon with an energy, $E > 2\Delta$, and it is unlikely that this phonon will escape the electrodes before it

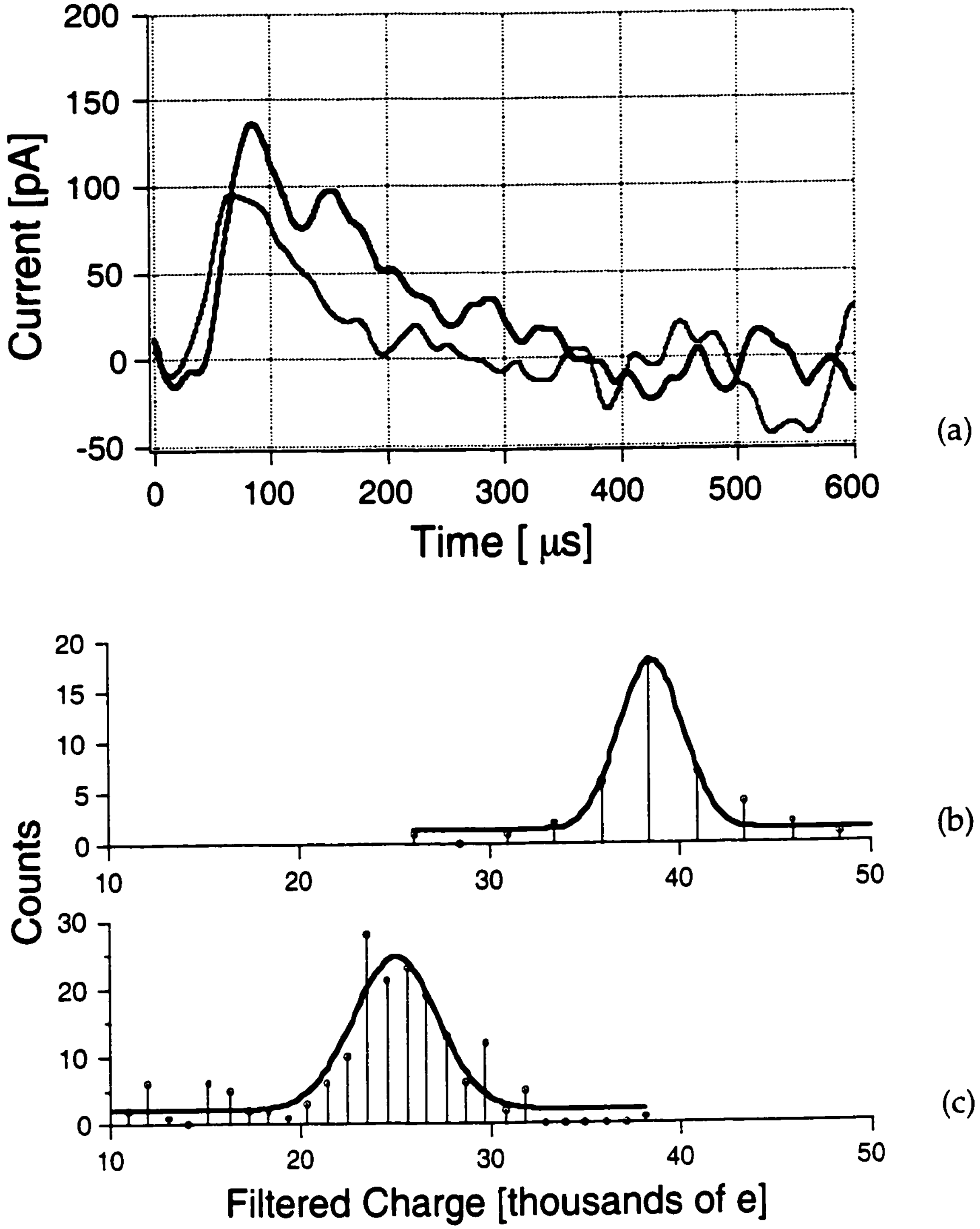


Figure 5.2 The top plot (a) shows current pulses from the two junctions of device OPS-F99-6A in response to the absorption of a single UV photon. We infer that the photon landed near the center of the devices. Plots (b) and (c) show histograms of collected charge in response to illumination by 4.89 eV UV photons and 2.27 eV green photons, respectively. The raw current pulses are digitally filtered before being integrated. The histograms represent photons absorbed in a limited region of the absorber.

breaks another pair. Thus, very little energy is removed from the quasiparticle system by recombination. On the other hand, when a quasiparticle is trapped into a region of normal metal, it will emit a phonon with an energy $E \sim \Delta$. This phonon cannot break a pair. Even if two quasiparticles are trapped, they simply emit two phonons that cannot break pairs. It is therefore possible that trapped fluxons in the electrodes could, in effect, "cool" the junctions. We do *not* have a detailed model that explains why the traps would confine quasiparticles generated in steady-state preferentially to photon-excited quasiparticles. We do expect, and have demonstrated at X-ray energies, that the photon-excited quasiparticles have a "hotter" energy distribution than the steady-state quasiparticles. It may be that the higher-energy, photon-excited quasiparticles are less likely to lose enough energy to scattering to become confined in the traps.

We have detected optical and ultraviolet photons using a number of different backtunneling devices. We have found the best resolution with the imaging device OPS-F99-6B. This device has a Ta absorber 10 μm wide by 100 μm long by 0.6 microns thick. Each Al trap overlaps the absorber by 5 μm. In Fig. 5.2a, we show a pair of current pulses which are the response of the detector to a 4.89 eV photon. Each trace is the response of one junction of the two junction device. We infer that this photon was absorbed near the center of the device because there are almost equal charges in the two current pulses. In Fig. 5.2b and 5.2c, we show two histograms of events recorded with this detector. We used a bias voltage of 60 μV for these measurements. Fig. 5.2b and Fig. 5.2c are the response to illumination with 4.89 eV ultraviolet photons and 2.27 eV green photons, respectively. We have plotted the number of events versus the collected charge. The raw current pulses were digitally filtered before being integrated to obtain the charge measurements.

The full width at half maximum (FWHM) of the histogram for 4.89 eV photons is 0.4 eV. The FWHM of the histogram for 2.27 eV photons is 0.47 eV. The histograms in Fig. 5.2 contain events from a limited range of the absorber. The theoretical limits based on the statistics of backtunneling are 0.2 eV at 4.89 eV and 0.13 eV at 2.27 eV. The limit calculated in the last chapter based on thermodynamic fluctuations of the steady-state quasiparticles is 0.45 eV, independent of energy. The measured resolutions are in good agreement with the thermodynamic prediction.

The resolution over the whole absorber, under UV illumination, is 1.5 eV, for a resolving power of R=3.5. An energy resolving power of R=3.5 in the UV implies that the detector can resolve 5 spatial pixels [Wilson 2000]. This particular detector has an active absorber area 70 μm long by 10 μm wide. Because the low signal-to-noise ratio, it was not possible to reliably correct the measured charge for loss as a function as position. Therefore, the degradation of the resolution when events from the whole absorber are included may be caused quasiparticle loss or by variations of the Ta gap energy near the Nb ground contact or near the region where the Al trap overlaps the Ta.

We have measured the effective quasiparticle recombination time by averaging 2000 single UV photon pulses and fitting an exponential decay to the waveform. We measure an effective quasiparticle recombination time of $\tau_{rec}=159\mu s$. The average number of times that a quasiparticle tunnels is $n=\tau_{rec}/\tau_{tun}$, where τ_{tun} is the tunnel time. We can extract the tunnel time from measurements of R_{NN}. We find $\tau_{tun}=2.46\ \mu s$. With this we find that n=65. We estimate that the initial number of quasiparticles created by a 4.89 eV photon is about $N_o=4000$ [Rando 1992]. The total number of electrons we collect is about $N_{total}=240{,}000$. Thus, the addition of the Ta plugs in the wiring does produce a large charge gain consistent with the observed recombination time.

Our measurements of backtunneling devices have shown that they do not provide good energy resolution for two main reasons. The first is that thermodynamic fluctuations of the quasiparticle gas in the junction electrodes lead to excess current noise. The other reason is that the size of these fluctuations is greatly increased by self-heating of the junctions, which increases the number of quasiparticles. We imagined two ways to move forward at this point. One was to remove the Ta plugs from the device designs, eliminating backtunneling altogether. A second approach would be to lower the gap of the quasiparticle plugs, by making a Ta/Al bilayer for instance. As we have said, we believe that the microscopic process that leads to excess current is quasiparticles tunneling multiple times without scattering, until they reach an energy above the gap greater than $2\Delta_{Al}$. The gap difference between pure Ta and Al is so large that even these high energy quasiparticles are confined in the junction electrodes. On the other hand, if the gap difference between the plug and the Al electrode is less than $2\ \Delta_{Al}$, the energetic quasiparticles may escape the junction before scattering. Still, if the gap difference is greater than eV at the typical bias voltage, there could still be a fair amount of backtunneling charge gain.

It is our interpretation that a related strategy has been pursued by the group at ESA, though they apparently came to their designs by empirical methods. The STJ detectors used by ESA are vertically stacked Ta/Al/Al-oxide/Al/Ta tunnel junctions. Their Al trapping layers are generally much thinner than the Ta layers. This causes the gap of the Al to be increased relative to the bulk value, due to the proximity effect. ESA systematically measured different devices with varying thicknesses of Al [Verhoeve 1997]. They found that the amount of backtunneling gain increased with increasing Al thickness. This is what we expect, since as the Al gets thicker, its gap decreases and the trap gets deeper. However, they also found that devices with thick Al showed excess current and excess noise. At the time, they had no explanation for this. They proceeded without an explanation, choosing as their standard design the Al thickness that offered the best balance between backtunneling gain and excess noise. In retrospect, it seems very likely that they were observing the same phenomena that we have described and explained.

5.3 Non-backtunneling Devices

We have decided to develop non-backtunneling devices for the optical/UV energy range. There are several reasons for doing this. First of all, the intrinsic energy resolution of non-backtunneling device is better than backtunneling devices. Second, non-backtunneling device are much faster. Ultimately, the speed of a non-backtunneling device is limited by the tunnel time of the device. The speed of a backtunneling devices is limited by the quasiparticle lifetime, which is much longer. (In fact, since the backtunneling gain is $n = \tau_{rec}/\tau_{tun}$, the detector is slowed by the same factor that the charge is amplified.) Finally, we understand the physics of non-backtunneling devices in great detail, based on many years of development in the X-ray energy range. The quasiparticle creation process may be very different in the optical than in the X-ray. However, we know that most processes that affect the resolution in our X-ray detectors occur in the tunnel junction [Segall 2000]. In addition, we expect that the difference in the creation processes will be unimportant by the time quasiparticles diffuse through the absorber to the tunnel junction. So, we expect that the most important physics issues will be the same in the X-ray and optical.

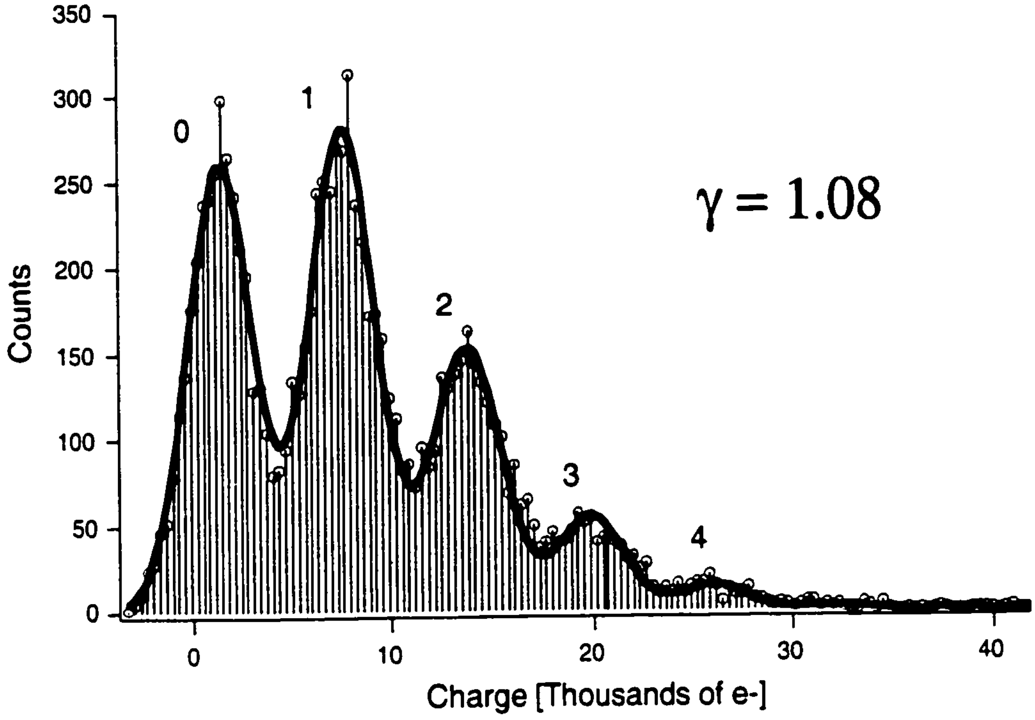
ESA has measured large variations of the responsivity as a function of energy in their STJ detectors [Kozorezov 2001]. Their explanation for this phenomenon is that there is a significant density of quasiparticle traps in their electrodes. At low energies, these traps trap a substantial fraction of the photon-induced quasiparticles, leading to a low responsivity. At higher energies, the traps are saturated by a small fraction of the quasiparticles and the responsivity increases. Comparing the optical/UV measurements and X-ray measurements at 6 keV, we do not see a significant variation in the responsivity of our detectors.

We have measured non-backtunneling devices using the same photon source, the mercury calibration lamp, with which we tested backtunneling devices. Unfortunately, the low electronic signal-to-noise makes it difficult to find an oscilloscope trigger level that captures all the photon pulses without also capturing many false triggers. This makes it difficult to accurately characterize the energy resolution of the device using this source. Still, we have been able to understand some device physics based on these measurements. We have measured two non-backtunneling devices, OPS-E00-7A and OPS-E00-10A. The device parameters are nominally the same except OPS-E00-10A is actually a backtunneling device that we converted by circumventing the Ta plugs with a new wiring layer. If we average a number of single photon pulses and fit an exponential to the tail, we find a fall time of about 7 μs for both devices. Overall, we expect the fall time to be limited by the tunnel time from the trap. The tunnel time for these samples is τ_{tun} = 4 μs for a bias voltage of 100 μV. One explanation for the longer fall time could be poor outdiffusion. If quasiparticles can diffuse freely in the counter electrode, it should only take about 10 ns for them to move away from the tunnel junction. However, if there is a barrier between the counter electrode and the wiring, the quasiparticles could be confined for a much longer time. If we put this assumption into our device simulation, we extract a 7 μs fall time, assuming an outdiffusion time from the counter electrode of τ_{out} = 0.1 μs. (In this case, it is really an out-tunneling time, but we will stay with the conventional name.) It is at first surprising that such a short out-diffusion time leads to a much longer fall time. We expect a long outdiffusion time to lengthen the pulse because quasiparticles confined in the counter electrode can backtunnel to the trap. If the out-diffusion time is longer than the tunnel time we expect that the fall time will become the outdiffusion time. In addition, we expect to see charge multiplication due to the backtunneling.

The explanation for the long fall time in our detectors is more subtle . When a quasiparticle first tunnels from the trap to the counter electrode, it will have relatively high energy and it is possible for it to reverse tunnel as an electron (not backtunnel as a hole) until it scatters inelastically and loses energy. In fact, since the quasiparticle would reverse tunnel to near the gap edge in the trap where the density of states is very high, the reverse tunnel time is very short compared to the backtunneling time. For our devices, we calculate that the reverse tunnel time to the gap edge is $\tau_{rev} = 0.1$ μs, compared to a backtunneling time of $\tau_{back} = 0.65$ μs. (The backtunneling time, τ_{back}, from the counter electrode is much shorter than τ_{tun} because the counter electrode volume is much smaller.) Still, this reverse tunneling process will only happen if quasiparticles do not scatter after tunneling. We calculate that the time for a quasiparticle 100 μV above the gap to scatter to any lower energy is 0.05 μs [Segall 2000]. This implies that, in fact, a significant fraction of quasiparticles reverse tunnel. Once they have reverse tunneled they wait, on average, another tunnel time in the trap. In this way a relatively short τ_{out} can lead to a lengthening of the pulse. We also note that this process does not lead to any charge multiplication.

We have been able to measure the energy resolution of non-backtunneling devices using the pulsed N_2 laser as a photon source. The N_2 laser emits intense pulses of light at 3.68 eV (337 nm). The length of the laser pulses is about 4 ns. We can use the laser as a multiphoton source by adjusting the intensity such that more than one photon is absorbed during each laser pulse. We believe multiphoton absorption can effectively simulate the response of the tunnel junction to a single higher energy photon. The photons are absorbed in a few nanoseconds but the excited quasiparticles diffuse to the junction over a period of a few microseconds, allowing the quasiparticle distribution to smooth before reaching the junction. In addition, the physical processes in the tunnel junction take place on microsecond timescales (even 100 μs timescales) and should effectively average over nanosecond variations in the absorption. In addition, since the device performance is limited by processes in the junction, multiphoton absorption should accurately simulate the energy resolution of the device.

The results in this section were obtained after processing the raw pulse data with a digital matched filter. The matched filter template is produced by averaging all the pulses in a data set to give an estimate of the ideal pulse shape. The filter template is then normalized. An individual pulse is then processed by first multiplying each pulse by the filter template. The filtered "charge" of each pulse is then calculated as the definite integral of the product waveform. We



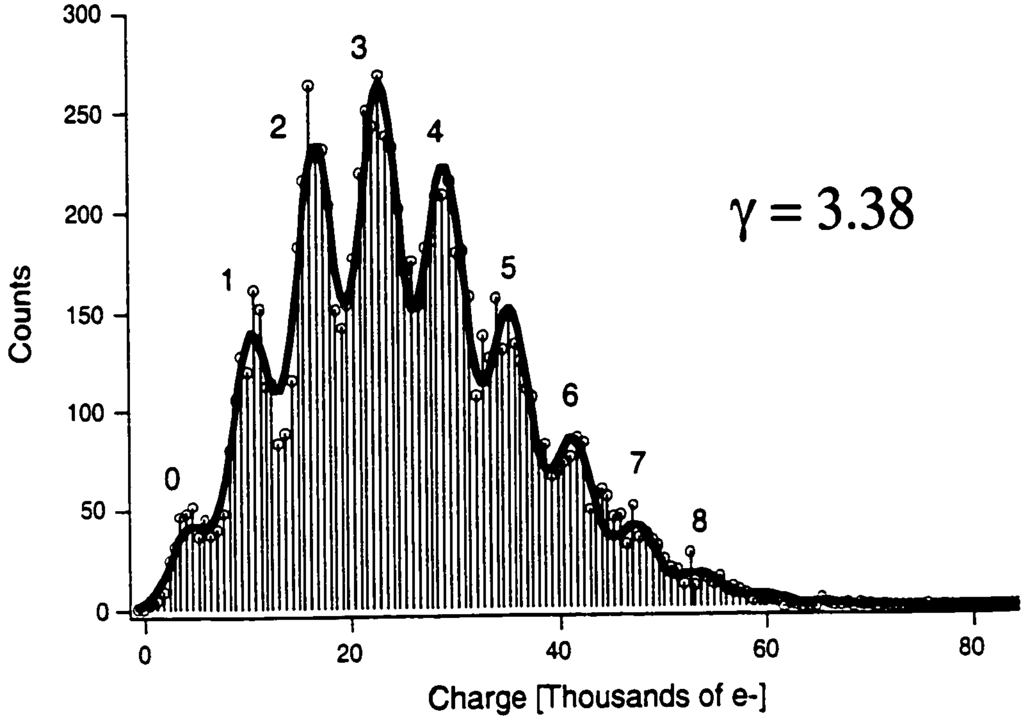


Figure 5.3 Histograms of the charge output of device OPS-E00-7A in response to illumination with the N_2 laser. The histogram is fit with a composite function that distributes the counts according to a Poisson distribution. The two histograms represent the response for two different laser intensities. Each graph is labeled with the average number of photons, γ, extracted from the fit. The peaks are labeled with the corresponding number of photons absorbed in the STJ.

choose to normalize the filter template such that if the template is applied to the average pulse, the output is equal to the *unfiltered* charge of the average pulse. In the special case that the noise of the system is purely white noise, the matched filter is an optimal filter [McDonough 1995]. We can also obtained similar results processing the data with a 5th order Chebyshev bandpass filter with a bandpass between 100 Hz and 35 kHz [Horowitz 1989]. We will use these numbers for the bandwidth of our measurements in our later discussion of noise.

We have measured a number of single junction devices, all with 100 μm^2 tunnel junctions. In Figure 5.3, we show two histograms of the charge output of device OPS-E00-7A under laser illumination. Each point in the histograms represents the response of the detector to a single pulse from the N_2 laser. The histograms shows a number of evenly spaced peaks, indicating that the number of photons absorbed varies from pulse to pulse. We expect this. We attenuate the laser's output at room temperature by 9 orders of magnitude, so that the probability of any one photon getting from the laser to the device is very small. If there is no correlation in the transmission of the photons, we expect the number of absorbed photons to follow a Poisson distribution. To test this hypothesis, we fit the histogram with a composite distribution:

$$P(Q)=\sum_{n=0}^{\infty}\left(\frac{\gamma^{n}e^{-n}}{n!}\right)\frac{1}{\sqrt{2\pi\sigma^{2}}}\exp\left(-\frac{1}{2\sigma^{2}}\left(Q-(R\bullet nE_{\gamma}+Q_{0})\right)^{2}\right)$$

where γ is the average number of photons absorbed, E_γ is the energy of a single photon, R is the responsively, Q_0 is the charge offset, and σ is the standard deviation of the charge. The composite distribution is the sum of many identical Gaussian distributions evenly spaced in charge. The amplitude of each Gaussian is determined by a Poisson distribution. The energy resolution is calculated from the fit as $\Delta E = 2.355\sigma/R$. The fits to the composite function are shown in Fig. 5.3 and we see that they agree quite well.

We can gain information by examining the structure of the histograms in more detail. First, we note that we get a better fit to the histograms if we assume the first peak represents laser pulses during which zero photons were absorbed.

This is reasonable because we use a synchronization signal from the laser to trigger the data acquisition by the oscilloscope. Because of this, we actually expect to record some triggers during which no photon was absorbed. Second, we notice that the average charge of the zero-photon peak is not zero. This charge offset, Q_0, is caused by phonons coupling from the substrate to the detector. As we have said earlier, the light emitted by the fiber onto the cold stage is dispersed over an area much larger than the detector. Thus, when the detector absorbs one photon, the substrate simultaneously absorbs ~10^4 photons. This energy is converted into phonons which can then couple to the detector and break pairs. Even though the coupling between the substrate and the detector is not very efficient, the number of quasiparticles created is significant because so much more energy is absorbed by the substrate than by the STJ.

We can distinguish this substrate signal from the aborber signal in the current pulses that we detect. In Fig. 5.4a we plot three waveforms, which were recorded from a second device, OPS-E00-10A. Each waveform was produced by averaging pulses whose charge fell within a single peak of a multipeak histogram. In this case, the waveforms are the average of pulses from the 0, 1, and 2 photon peaks of a histogram with $\gamma = 0.80$ photons. The pulse from the 0 photon peak has a uniform decay with a very long time constant. We see that the 1 and 2 photon pulse shapes instead have a double exponential decay. They both have a fast component whose magnitude scales with the number of photons absorbed in the STJ. In addition, the time constant of the fast component matches the time constant measured when the device is excited with single photons from the Hg lamp. We also clearly see that the slow components from the 1 and 2 photon pulses are the same and they also match the decay of the 0 photon pulse. Therefore, we conclude that the slow component is independent of the number of photons absorbed in the STJ and we infer that is due to excitation of the substrate. In Fig. 5.4b we plot three more waveforms. These waveforms are the average pulse from the 0 photon peak of three different data sets, each recorded with different laser intensities. The average number of

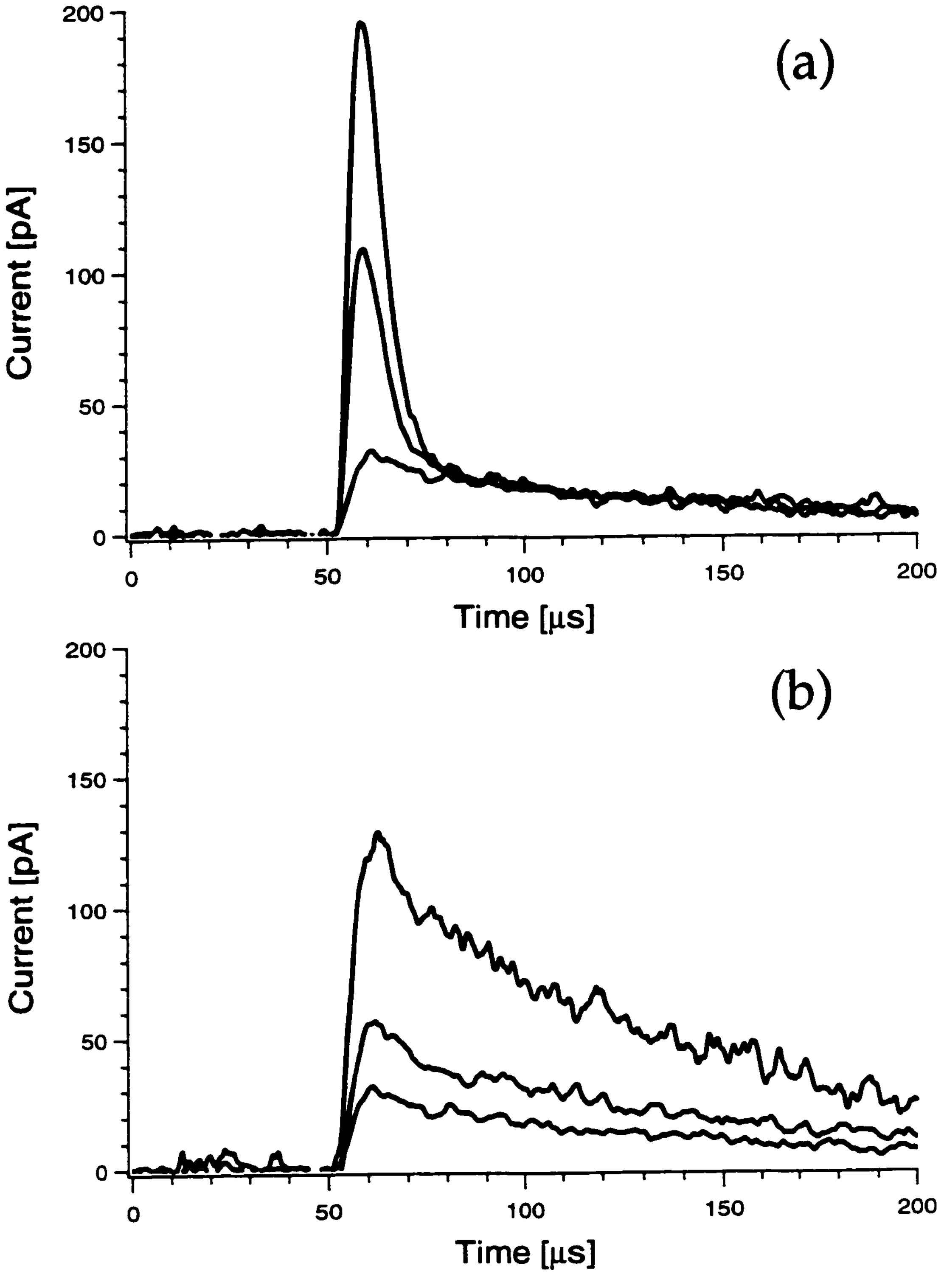


Figure 5.4 Average current pulses from device OPS-E00-10A. The pulse shapes are produced by averaging pulses whose charges are in one peak of a charge histogram (see Fig. 5.3). In (a), the pulses are, in order of increasing peak current, from the 0, 1, and 2 photon peaks of a histogram with $\gamma = 0.80$. We see the pulses have a fast component that scales with the number of photons and a slow component that is the same for all three. In (b), the pulses are from the 0 photon peak of three different histograms with, in order of increasing peak current, $\gamma = 0.80$, 1.56, and 3.65.

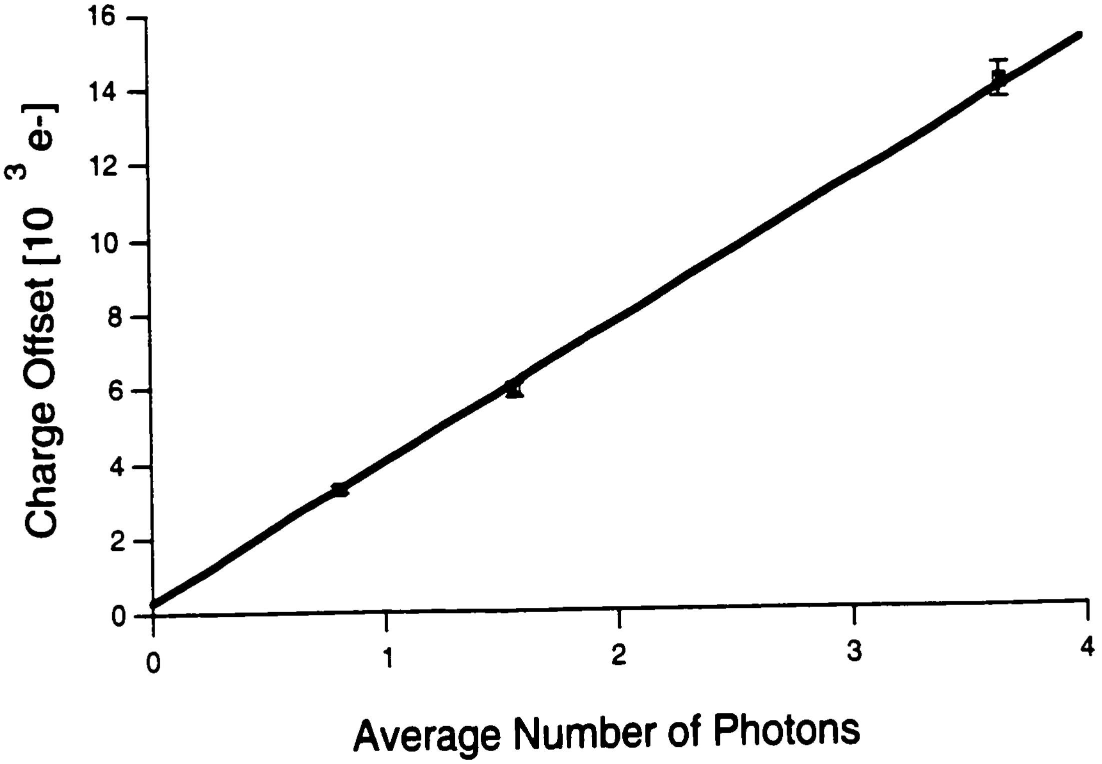


Figure 5.5 Variation of the charge offset of the zero photon peak as a function of the average number of photons absorbed in the detector. The average number is varied by varying the room temperature attenuation of the laser.

photons absorbed for the three data sets were $\gamma = 0.80, 1.56$, and 3.65. (The pulse for $\gamma = 0.80$ is the same as in Fig. 5.4a.) We see that the magnitude of the substrate signal is proportional to the average number of photons, which is a measure of the laser intensity.

We exclude much of the substrate signal simple by windowing the pulses in time. Typically, we only use a 30 μs section of the pulses. Still, we expect the substrate effect to produce a charge offset in the histograms because some of the substrate signal is coincident in time with the absorber signal. As we have seen, this substrate signal is independent of the number of photons absorbed in the STJ. This implies all the peaks should be shifted to larger charge by the same average amount. Our composite fitting function therefore assumes that the peaks are equally spaced, but have an offset from zero. This assumption clearly fits the data. We do expect that the charge offset will vary with the attenuation of the laser intensity. If we decrease the attenuation, the substrate will absorb more

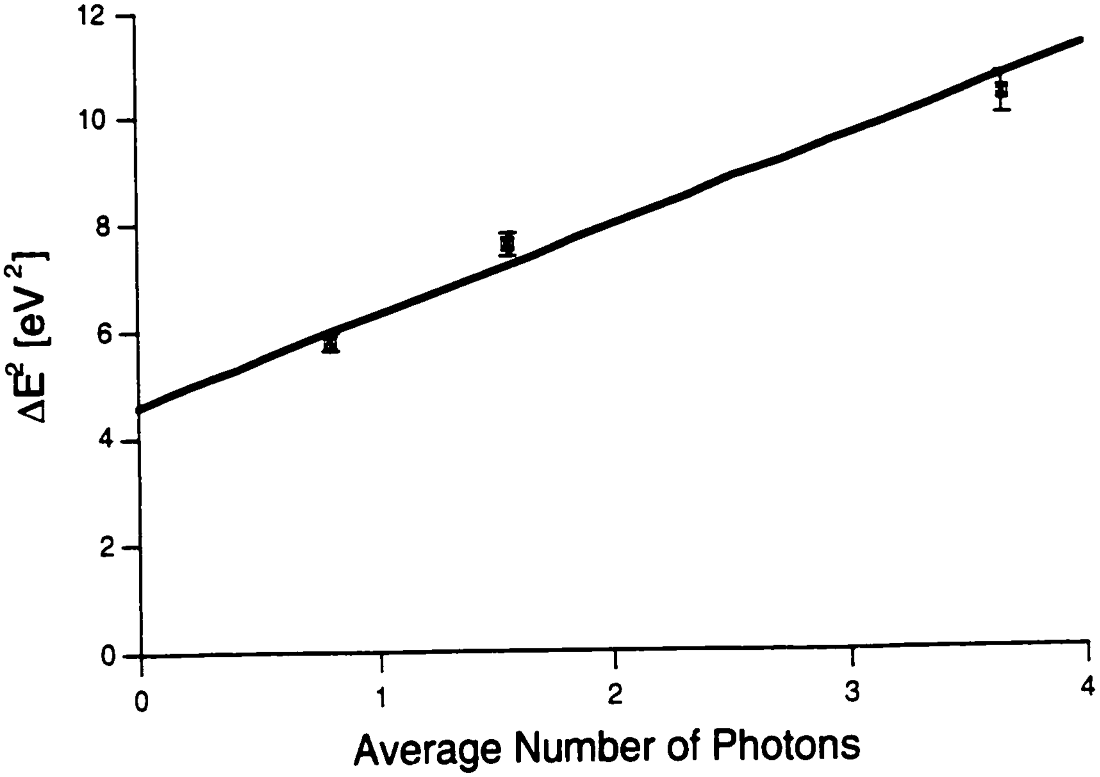


Figure 5.6 The square of the energy resoltuion plotted against the average number of absorbed photons. The y-intercept gives the resolution not due to substrate noise.

energy during each pulse, ultimately breaking more pairs. Of course, decreasing the attenuation will also increase the average number of photons absorbed by the STJ. Figure 5.5 shows the variation in the charge offset as a function of the average number of photons absorbed. These offsets are obtained by fitting the histogram of charges, after windowing and filtering. (The average pulses in Fig. 5.5 were taken from the same three data sets.) We see that the offset varies linearly with the average number. We also see that the intercept is only slightly different from zero, meaning that if we turned the laser intensity to zero, there would be a negligible offset.

We also expect the substrate charge to degrade the energy resolution of the detector. This is because the amount of charge created by substrate phonons will vary from pulse to pulse. We can infer what part of the measured resolution is due to this substrate noise by studying how the resolution changes when we change the attenuation. To do this, we must make some assumptions. First, we assume that the substrate noise is independent and adds in quadrature with other noise sources. We can also guess that the variance of the substrate charge will be proportional to the average substrate charge. In Fig. 5.6 we show the square of the energy resolution as a function of the average number of photons absorbed in the STJ (which is proportional to the offset charge). We observe a number of things. First of all, the energy resolution clearly broadens as the average number of photons increases. It is also clear that for the lowest number, the energy width is not dominated by the substrate noise. If we fit a line to the data, we find that the substrate noise is about 1.3 eV per photon absorbed on average in the Ta and the remaining noise is about 2.14 +/- 0.07 eV.

We now want to understand what factors contribute to this 2.14 eV energy width. Electronic noise is certainly one important contribution. We estimate the broadening due to the electronic noise in the following way. During a run, we can shutter the N_2 laser while still allowing it to electronically trigger the oscilloscope. We can then process these noise pulses as if they were real photon pulses. If we do this, we get a histogram of charges with zero mean and some variance. Based on the responsivity measured from the multiphoton histogram, we can convert this charge noise to an energy resolution. We find an electronic noise resolution of 2.15 eV. Thus, within our measurement accuracy the resolution is fully accounted for by electronic noise and substrate noise.

5.4 Future Work

Future work on this project will roughly fall into two categories: improving the performance of the detectors themselves and developing applications. I will focus on what can be done to improve the performance of the detectors here. At the moment, the performance of the detectors is limited by the electronic noise of the amplifier and the noise due to substrate absorption. The substrate noise is essentially an artifact of using the N_2 laser as a source with inefficient coupling. It is not a characteristic of the STJ detector itself. We would

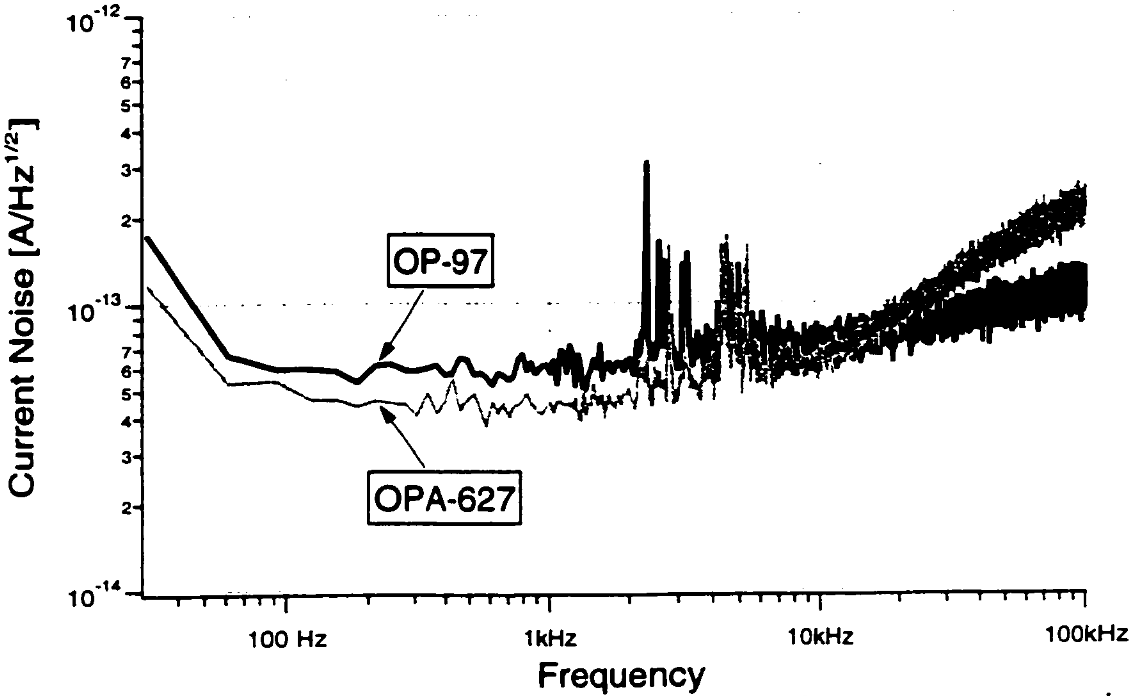


Figure 5.7 Current noise spectra of the detector plus amplifier. We plot two spectra corresponding to two different nulling loop amplifiers, the OP-97 and OPA-627.

not observe the same noise if we measured with a single photon source. To use the N_2 laser more effectively in the future, we could either improve the coupling of the fiber to the detector, or could illuminate the detector through an aperture, masking the substrate from the laser light. Thus, improvements in the device performance in the near future will mostly come improvements in the electronic noise.

There are four major sources of electronic noise in our existing measurement setup. They are electromagnetic interference, current noise of the tunnel junction, voltage noise of the input FET and current noise of the nulling loop voltage follower. Referring to Fig. 5.7, we see a group of discrete lines in the electronic noise spectrum in the frequency range from 1-10 kHz. It is not clear whether these lines are due to direct coupling of electromagnetic radiation, or if they are generated by microphonic vibrations. Either explanation is feasible. The attenuation of the shielded enclosure we use drops off below 10 kHz, allowing low-frequency magnetic fields to penetrate. If this is the cause, we could reduce the effect by using a μ-metal shield that effectively attenuates low frequency magnetic fields. We could also improve the wiring of the dewar to reduce its susceptibility to EMI. If the noise is coupled by microphonic vibrations, we could improve the dewar's vibration isolation (which is now very crude) and improve the wiring.

We naively expect the current noise of the junction to be the shot noise of its bias current. This implies a current spectral density of $i_n = (2eI)^{1/2}$. A bias current of approximately 1 nA would then imply $i_n \sim 20$ fA/√Hz. We measure a current noise from the junction of $i_n \sim 40$ fA/√Hz. There are a number of things that could explain this excess noise. First of all, the noise could be caused by a flux of infrared or microwave photons being absorbed by the device. As mentioned in Chapter 3, the current generated by a flux of photons from room temperature blackbody radiation would have a noise greatly exceeding standard shot noise. Thus, even if this photon flux were generating a negligible amount of dc current, its noise could dominate the shot noise of the device. At the moment, the wiring in the dewar is not filtered, so black body radiation from the N_2 stage and room temperature could couple down the leads to the device. If this is the problem, the solution is to filter the leads and enclose the detector in a light-tight enclosure at 0.2 K.

We believe there are also intrinsic physical processes that could lead to excess current noise in the device. First of all, the standard derivation of shot noise in a tunnel junction assumes that the electrodes are ideal electron reservoirs, meaning that the tunneling does not affect the quasiparticle distribution in the two electrodes. However, at least one of our electrodes, namely the trap, is finite in extent. We therefore expect that the number of quasiparticles in the electrode will fluctuate of order the square root of the number. These fluctuations will in turn generate excess current noise. The calculations and simulation presented in the previous chapter show that the shot noise power is doubled for frequencies less than $\sim 1/\tau_{tun}$. As mentioned earlier in this chapter, we have evidence that a significant number of quasiparticles that tunnel across the junction reverse tunnel as electrons before they cool to the gap edge. This is distinct from backtunneling, where quasiparticles tunnel as holes. This combination of tunneling and reverse tunneling contributes no average current, but still contributes shot noise. Basically, the dc current observed should be $I_{dc} = |I_{forward}| - |I_{reverse}|$, while the shot noise power is $2eI_{tot}$, where $I_{tot} = |I_{forward}| + |I_{reverse}|$. Thus, we see that a significant amount of reverse current would cause us to underestimate the shot noise based on the observed dc current. For either of these explanations, the solution is to reduce the number of steady-state quasiparticles in the junction, either by cooling the junctions or making smaller junctions.

It is not practical to reduce the voltage noise of the input FET. The FET we presently use, the 2SK146, has the lowest voltage noise of any commercial FET. (The noise is specified to be 0.6 nV/√Hz. We actually have a number of FETs with a voltage noise of 0.4 nV/√Hz.) However, the contribution of the voltage noise to the total noise depends on the impedance at the input of the amplifier. In fact, the induced current noise due to the voltage noise, e_n, is

$$i_n = \frac{e_n}{Z_{in}}.$$

Thus, we can reduce the contribution of the voltage noise by increasing the impedance at the input. At the moment, the impedance at the input is dominated by the capacitance of the 2SK146 and the capacitance of the leads running from room temperature to the 0.2 K stage. The effective capacitance of the 2SK146 when used with the A250 is C_{FET} = 300 pF, including about 100 pF arising from the Miller effect. The capacitance of our leads is C_{lead} = 120 pF. The best overall strategy for reducing the lead capacitance is to move the input FET from room temperature to the cold stage. Unfortunately, silicon JFET's do not function properly below 120-150 K. Therefore, we would place the FET as close to the cold stage as possible within the constraints of the cooling power of dewar and the practicality of thermal engineering. Some groups [Porter 1999,2000] have engineered designs with as little as 5-7 pF of capacitance between 150 K FETs and 0.05 K detectors! We believe that 20-30 pF might be a reasonable projection for our dewar. In addition, not every FET can even be cooled to 150 K. We have found that the 2SK146 does not function properly at cryogenic temperatures. We have recently made measurements of the IF1801 FET made by InterFET. We found that it does function well at cryogenic temperatures and that its noise improves from 0.7 nV/√Hz to 0.5 nV/√Hz upon cooling. Its capacitance is specified to be 100 pF. This is a good candidate for a cold FET. InterFET makes other FETs that would be better matched to 20-30 pF of lead capacitance, but it remains to be seen if they function properly at cryogenic temperatures.

The final source of electronic noise in our amplifier is the current noise of the input operation amplifier (op-amp) of the nulling loop, which provides the dc bias path of the amplifier. In the electronics schematic of Fig. 2.3, this amplifier is labeled "OP-97". We presently use the OP-97 op-amp, which is a precision bipolar transistor (BJT) op-amp designed to have relatively low current noise. We measure a white current noise of 40 fA/√Hz. While this is much less current noise than typical BJT amplifiers, it is a dominant contribution to our noise at low frequency, and represents a lower resolution limit of about 1 eV. We have investigated using op-amps with FET inputs. FET op-amps have extremely low current noises at low frequency, typically ~ 1 fA/√Hz due to the shot noise of their input gate leakage. Unfortunately, FETs have another source of current noise. At higher frequencies, when there is a phase shift at the drain of the input FET, the Miller effect no longer leads to an effective input capacitance but , instead, produces an effective input resistance. This resistance has Johnson noise. Further, this Miller resistance decreases linearly with frequency, producing a current noise that increases as the square root of frequency [Horowitz 1989]. This noise typically starts to dominate the shot noise of the FET in the frequency range of 10-100 kHz. We have measured the current noise of a number of precision FET op-amps (including the AD-795 and OPA-627) and found that any benefit gained from the decrease in the low frequency noise (compared to the OP-97) is canceled by the increasing Johnson noise at high frequency (refer to Fig. 5.7).

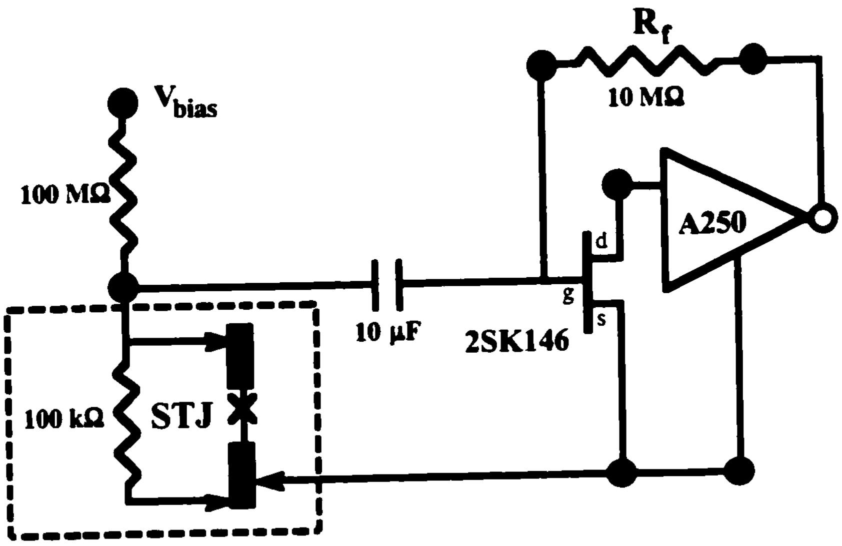


Figure 5.8 Schematic of passive voltage bias circuit, showing a cold bias resistor and AC-coupling capacitor.

Our solution to this problem is to move from the active voltage bias circuit, which uses the op-amp, to a passive voltage bias. In Fig. 5.8, we show the schematic of a passive voltage bias circuit that we have tested. The are a number of reasons why we have used the active voltage bias instead of a passive voltage bias in the past. Basically, a passive voltage bias circuit with the same dc load line as the active circuit would contribute many orders of magnitude more noise. For instance, a conservative estimate of the dc load line of our active circuit is 10 Ω. A *passive* 10 Ω bias would also appear as 10 Ω at the input of the 2SK146/A250 amplifier. This 10 Ω impedance would convert the voltage noise of the 2SK146 to 40 pA/√Hz, which is 1000 times higher than the noise of the active bias circuit! Still, the dc stability requirements that the bias circuit must meet are much less stringent for optical detectors than for X-ray detectors, especially for a single pixel device. Basically, a bias resistor between 100 kΩ and 1 MΩ cooled to 0.3 K should be sufficient from the point of view of noise. Our present optical/UV STJs have a dc bias current of ~ 1 nA at 100 µV. This implies an effective dc resistance of $R_{dc} = V_{dc}/I_{dc}$ = 100 kΩ. So, a 100 kΩ bias resistor is a marginal choice, remembering the load line of a good voltage bias should be much less than the R_{dc} of the device. In fact, we have measured devices with the circuit shown in

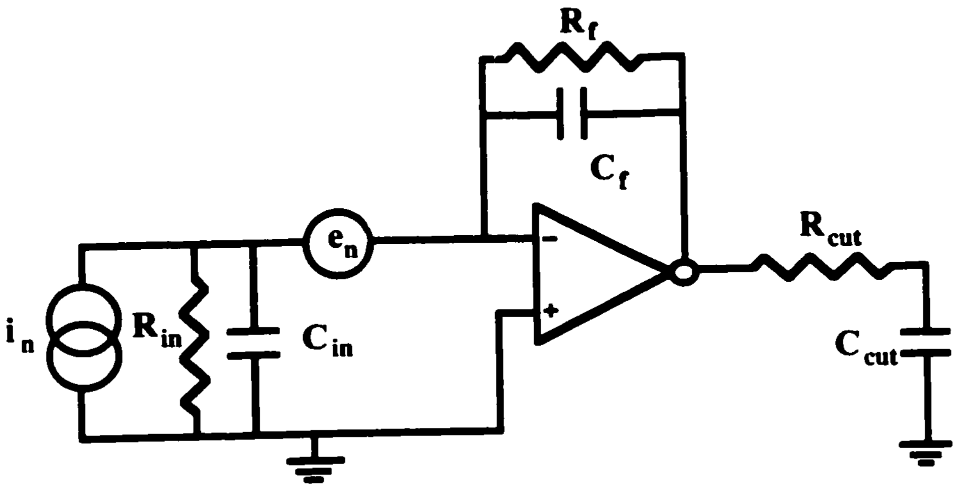


Figure 5.9 Conceptual circuit used to calculate how the noise of the measurement system scales with various parameters. The circuit is comprised of an integrator amplifier followed by a simple filter.

Fig. 5.8 and found exactly that: the bias was marginally stable. Still, by either making smaller junctions or cooling the existing junctions, we can make R_{dc} much higher. Basically, the voltage scale is fixed by Δ_{Al} but the current decreases linearly with junction size and exponentially with decreasing temperature. Thus, we believe that passive voltage bias will work well in the near future.

We can predict how the improvement mentioned above will affect the energy resolution of the detectors. To do this, we must understand how the energy resolution scales with various parameters that affect the noise. For the purpose of these calculations, we imagine the conceptual measurement circuit show in Fig. 5.9. The first part of the circuit is an integrating amplifier, characterized by an integration time $\tau_f = R_fC_f$. The integrator is followed by a simple filter that limits the bandwidth of the measurement. The filter is characterized by the time constant $\tau_{cut} = R_{cut}C_{cut}$. Without loss of generality, we can express the time constants of the circuit in units of the τ_{tun}, i.e., $\tau_f = \alpha\tau_{tun}$ and $\tau_{cut} = \tau_{tun}/\beta$, where we general expect that $1 < \alpha,\beta < 10$. Expressing the time constants of the measurement in terms of τ_{tun} will allow us to more clearly understand how the noise performance is related to the speed of the detector. In the end, we want to compute the signal-to-noise ratio at the output of this circuit.

However, for an input charge signal Q_γ, the output voltage will be $V_{out} = Q_\gamma/C_f$ within factors of order unity for a wide range of parameters. For that reason, we will focus on how the noise scales.

In general, we can compute the variance of the charge noise referred to the input, σ_q^2, as

$$\sigma_q^2 = \frac{C_f^2}{2\pi} \int_0^\infty S_V(\omega)d\omega \tag{5.4.1}$$

where $S_V(\omega)$ is the single-sided, voltage-noise spectral density at the output. For an input white current-noise spectral density i_n^2, we compute

$$\begin{aligned} \sigma_{q,i}^2 &= \frac{i_n^2}{2\pi} \int_0^\infty \frac{\omega_{cut}^2}{\left(\omega^2 + \omega_{cut}^2\right)\left(\omega^2 + \omega_f^2\right)} d\omega \\ &= \frac{i_n^2}{4} \frac{\omega_{cut}}{\omega_f\left(\omega_f + \omega_{cut}\right)} \end{aligned} \tag{5.4.2}$$

where $\omega_f = 1/\tau_f$ and $\omega_{cut} = 1/\tau_{cut}$. In terms of τ_{tun}, we then find:

$$\sigma_{q,i} = \frac{1}{2} i_n \sqrt{\tau_{tun}} \left(\frac{\alpha^2}{\alpha + {}^1\!/_\beta} \right)^{1/2} . \tag{5.4.3}$$

Thus, we see that the charge noise due to current noise is proportional to the magnitude of the noise and to the square root of the tunnel time. This implies that if the electronic noise is dominated by white current noise, then the signal-to-noise ratio will be maximized by making the detector as fast as possible.

We can now consider the contribution of an input white voltage-noise spectral density e_n^2. As mentioned earlier, the contribution of the voltage noise to the total noise depends on the impedance at the input of the amplifier. For an total resistance R_{in} and input capacitance C_{in}, we find

$$\begin{aligned} \sigma_{q,e}^2 &= \frac{C_{in}^2 e_n^2}{2\pi} \int_0^\infty \frac{\omega_{cut}^2\left(\omega^2 + \omega_{in}^2\right)}{\left(\omega^2 + \omega_{cut}^2\right)\left(\omega^2 + \omega_f^2\right)} d\omega \\ &= \frac{C_{in}^2 e_n^2}{4} \frac{\omega_{cut}\left(\omega_f \omega_{cut} + \omega_{in}^2\right)}{\omega_f\left(\omega_f + \omega_{cut}\right)} \end{aligned} \tag{5.4.4}$$

where $\omega_{in} = 1/R_{in}C_{in}$. In this case, we have not expressed τ_{in} in terms of τ_{tun} because we cannot freely choose R_{in} and C_{in} in the same way that we can choose τ_f and τ_{cut}. We can simplify (5.4.4) in two important limits. In the limit that R_{in} dominates the impedance in the signal band we have $\omega_{in}^2 >> \omega_f\omega_{cut}$. We then find

$$\sigma_{q,e}^2\left(\omega_{in}^2 >> \omega_f\omega_{cut}\right) = \frac{1}{4}\left(\frac{e_n}{R_{in}}\right)^2 \frac{\omega_{cut}}{\omega_f\left(\omega_f + \omega_{cut}\right)} \qquad (5.4.5)$$

which is the same as (5.4.2) with an effective current noise $i_n = e_n/R_{in}$. The opposite limit is when C_{in} dominates the input impedance, which implies $\omega_{in}^2 << \omega_f\omega_{cut}$. In this limit, we find that charge noise referred to the input is

$$\sigma_{q,e} = \frac{1}{2}\frac{C_{in}e_n}{\sqrt{\tau_{tun}}}\left(\frac{\beta^2}{\beta + \frac{1}{\alpha}}\right)^{1/2}. \qquad (5.4.6)$$

We see in this limit that the charge noise is proportional to magnitude of the voltage noise and input capacitance, but *inversely* proportional to the square root of the tunnel time. This implies that if the noise of the amplifier is dominated by C_{in} and e_n, then the detector should be made as slow as possible.

We can use the above formulas to scale the noise of our detector system. Still, we need a absolute calibration of the noise. The formulas above specify charge noise. However, we measure the effective current noise of our amplifier. In principle, it is possible to calculate the charge noise given a measured current noise spectrum, but, in practice, it is difficult to do in a way that can be compared absolutely to an output charge signal. In addition, typical specification for amplifiers and transistors are in terms of voltage noise and current noise. For these reasons, we calibrate the noise in terms of the measured effective current noise in the follow way. The average current noise of the noise spectrum in Fig. 5.6 in the band 100 Hz to 35 kHz is 85 fA/√Hz. Using 2.15 eV as the energy resolution due to electronic noise we obtain a conversion factor between effective current noise and energy resolution of 0.025 eV/(fA/√Hz). For a given set of noise parameters (e_n, i_n, C_{in}, etc.) we calculate the average effective current noise, $\bar{i}_n$, as

$$\bar{i}_n^2 = \frac{1}{f_{cut}} \int_0^{f_{cut}} \left[\left(\frac{e_n}{Z_{in}} \right)^2 + i_n^2 \right] df \tag{5.4.7}$$

where the cut-off frequency is

$$f_{cut} = 35\ \text{kHz} \cdot \frac{4\ \mu s}{\tau_{tun}}. \tag{5.4.8}$$

We then calculate the energy width due to the electronic noise as

$$\Delta E_{elec} = 0.025\ \text{eV} \cdot \bar{i}_n \cdot \sqrt{\frac{\tau_{tun}}{4\ \mu s}} \tag{5.4.9}$$

with $\bar{i}_n$ in units of fA/√Hz. The tunnel time is scaled with respect to the value $\tau_{tun} = 4\ \mu s$ because that is value for the detector measured, when biased at 100 μV. Conceptually, we are calculating an RMS current noise and then multiplying it by a factor that accounts for integrating the current noise in time. It is easy to show that this method reproduces the scaling properties calculated in (5.4.1)-(5.4.6).

In previous work at X-ray energies, we found a conversion factor of 0.072 eV / (fA / √Hz) for a double junction device. For a single junction device, this value should be divided by √2, giving 0.051 eV / (fA / √Hz). We believe the additional factor of 2 difference is related to improvements in grounding and signal processing. In older measurements, there was a significant amount of 60 Hz noise. Consequently, the digital filter that gave the best energy resolution typically had a low cut-off frequency in the 1-5 kHz range. While this filter effectively removed the noise, it also filtered a significant amount of signal, typically reducing the collected charge by a factor of 2-3. In recent measurements, the 60 Hz pickup has been eliminated, so we typically use a filter with a low cut-off frequency of 100 Hz. With this filter, the charge is *not* significantly reduced. Thus, the signal-to-noise is better and the conversion factor is smaller.

We plot predicted resolving powers for various scenarios in Fig. 5.10. The upper limit on the resolving power is the statistical limit based on creation noise, with a Fano factor of F = 0.2, and trapping noise, assuming a trapping multiplication of 2 [Segall 2000]. The total effective Fano factor, including the trapping noise, is F′ = 0.45. The energy width due to these statistical sources is

where $\varepsilon = 1.7\Delta_{Ta}$. All three curves assume τ_{tun}=10 μs, area = 5 μm^2, and passive voltage bias. We also assume that the junctions are operated at 150 mK and that their subgap currents follow BCS predictions down to that temperature. We included that shot noise of the bias current in the calculations, but it is unimportant at 150 mK. We plot curves for three different amplifiers. The first is a cold IF1801 FET with e_n = 0.5 nV/√Hz, C_{FET}=100 pF, C_{lead} = 20pF and R_{bias} = 10 MΩ. We specify the passive voltage bias resistance because the noise at low frequency is dominated by e_n/R_{bias}. The Johnson noise of the bias resistor is also included in the calculations, assuming the resistor is also cooled to 150 mK. The second amplifier uses values for cold FET amplifiers assembled by the Goddard/ Wisconsin collaboration [Szymkowiak]. The values are e_n = 3 nV/√Hz, C_{FET}+C_{lead} = 13 pF and R_{bias} = 10 MΩ. We believe that these values represent the limit of cold FET technology. The final curve is for a single-electron transistor amplifier [Schoelkopf 2001]. The parameters assumed are e_n = 20 nV/√Hz, C_{SET}+C_{lead} = 0.0 pF, C_{STJ} = 0.25pF and R_{bias} = 20 MΩ. The value of e_n assumed is better than what has been demonstrated (30 nV/√Hz), but is a reasonable estimate for future work. We assume a higher bias resistance for the SET to compensate for the higher value of e_n. We also assume that the SET is fabricated on-chip with the STJ such that capacitance of the SET and leads are negligible. The dominant capacitance in this case is the capacitance of the tunnel junction itself. In table 4.1, we summarize the electronic noise contributions for the various amplifiers. These contributions are independent of energy, except for the SET amplier whose total electronic noise depends on energy, as discussed below.

The only other important noise consideration is the reduction in the dynamic resistance of the tunnel junction during a photon pulse. The dynamic resistance during a pulse, R_d, is reduced because the effective temperature of the

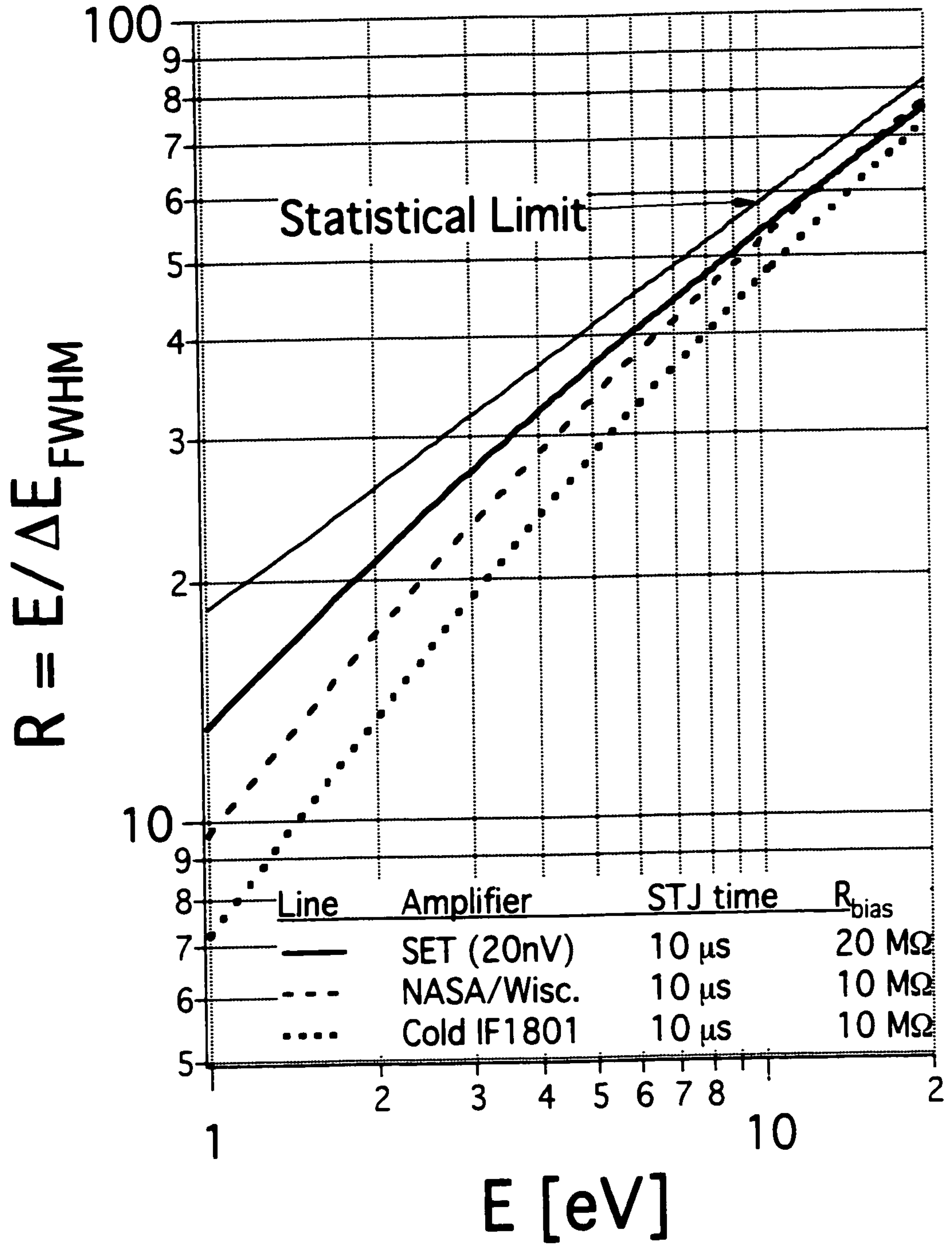


Figure 5.10 Predictions of resolving power for a single pixel STJ with a 5 μm^2 tunnel junction. The resolving power is bounded by the statistical limit imposed by quasiparticle creation trapping multiplication. We have plotted predictions for three different amplifiers.

Amplifier	**Noise due to $e_nC_{in}\omega$**	**Noise due to e_n/R_{bias}**	**Total Electronic**
Cold IF1801	0.12 eV	0.002 eV	0.12 eV
NASA/Wisc.	0.08 eV	0.012 eV	0.08 eV
SET	0.01 eV	0.04 eV	0.064 eV

Table 5.1 Predicted electronic noise contributions of the three proposed amplifiers. The capacitive noise is calculated by assuming the input resistance is infinite. The resistive noise is calculated by assuming the input capacitance is zero. Note that the two contributions are not completely independent when calculated in this way, because both terms are related to the amplifier voltage noise. The Johnson noise of the bias resistor contributes 0.036 eV to the first two amplifiers and 0.025 eV to the SET amplifier. The total electronic noise quoted for the SET is the value at 5 eV.

excited quasiparticles in the junction is generally higher than the bath temperature. For an incident photon energy of 6 keV, we have measured a resistance during a pulse of 1.7 kΩ for a junction bias voltage of 80-90 μV and τ_{tun} = 2.5 μs [Segall 2000]. We assume that R_d is inversely proportional to the photon energy, E_γ, simply because the magnitude of the current scales with the photon energy without changing the energy distribution of the quasiparticles in junction. We assume that the junctions are baised at 150 μV which doubles R_d. Changing τ_{tun} affects R_d in two ways. First, the magnitude of the excited current, and therefore R_d, is inversely proportional to τ_{tun}. Second, a longer τ_{tun} allows the quasiparticles to cool more before tunneling, increasing R_d. We have simulated this effect, and use a linearized fit to the simulations in the noise calculations. Including all of these effects, we calculate the resistance during a pulse as

. (5.4.10)

The effect of R_d is only important in the SET curve here. It is this effect that causes the SET curve to bend down from the statistical noise at high energy.

These predictions say that, with hard work, STJs should be able to achieve a an energy resolution that will make them useful for a number of applications.

5.5 Perspectives

Cryogenic detectors are developing into a mature technology. At this point, the important physics of both STJ detectors and TES detectors have been elaborated. Detectors in the laboratory have been shown to perform at or near theoretical expectations. In addition, prototype spectroscopy systems have been used to make real measurements. In short, the detectors themselves have developed to the point that they can start being used in real applications. The next challenge for people working in the field will be to develop truly useful systems that enable measurements of new physical phenomena. Especially in the optical energy range, the technical challenges are intense. Many optical applications will require large-scale, imaging arrays of detectors. It is not clear that it will be possible to develop measurement systems can that cool these large arrays while coupling light to them and bringing their output signals to room temperature without degrading the performance of the detectors. This is the next great challenge for the field, and the success or failure of this endeavor will determine the ultimate utility of cryogenic detectors. Depending on one's outlook, this challenge can be seen as daunting or exciting.

References

Booth, N. and D. J. Goldie, *Supercond. Sci. & Tech* **9**, 493 (1996).

Blanter, Ya. M. and M. Buttiker, *Phys. Reports* **336**, 1 (2000).

de Bruijne, J. H. J. et al., *Astronomy & Astrophysics* **381**, L57-60 (2002).

Cabrera, B. et al., "Detection of single infrared, optical, and ultraviolet photons using superconducting transition edge sensors", *Appl. Phys. Lett.*, vol. 73, pp. 735-737, 1998.

Cabrera, B. et al., in *Proc. 9th Int. Work. on Low Temp. Detectors*, edited by F.S. Porter et al., pp. 565-570 (AIP Conference Proceedings, vol. **605**, New York, 2002)

Chi, C.C., and J. Clarke, *Physical Review B* **19**, 4495 (1979).

Devoret, M., unpublished.

Face, D.W. and D.E. Prober, *Journal of Vacuum Science & Technology A* **5**, p.3408-11 (1987).

Friedrich, S., PhD Dissertation, Yale University (1997a).

Goldie, D.J. et al., *Applied Physics Letters* **64**, p.3169-71 (1994).

le Grand, J. B. et al., *Proc. 7th Int. Work. on Low Temp. Detectors*, p. 106 (1997).

Gray, K. E., *Appl. Phys. Lett.* **32**, 392 (1978);

Gray, K. E., *J. Phys. F: Metal Phys.* **1**, 290 (1971)

Gray, K. E., in *Nonequilibrium Superconductivity, Phonons, and Kapitza Boundaries*, edited by K. E. Gray, NATO ASI, Ser. B, Vol. 65 (Plenum, New York, 1981).

Henke, B. L. et al., *Atomic Data and Nuclear Data Tables* **54**, 181-342 (1993).

Horowitz and Hill, *The Art of Electronics* (Cambridge University Press, Cambridge, 1989).

Imry, Y., *Introduction to mesoscopic physics* (Oxford University, New York, 1997).

Kaplan, S. B. et al., *Phys. Rev. B* **14,** 4854 (1976).

de Korte, P.A.J. et al., *Proc. SPIE* **1743,** 24 (1992).

Kozorezov, A. G. et al., in *Proc. 9th Int. Work. on Low Temp. Detectors,* edited by F.S. Porter et al., pp. 51-54 (AIP Conference Proceedings, vol. **605,** New York, 2002).

Kozorezov, A. G. et al., *Applied Physics Letters* **78,** 3654-6 (2001).

Kraus, H. et al., "Quasiparticle trapping in a superconductive detector system exhibiting high energy and position resolution", *Phys. Lett. B,* vol. 231, pp. 195-202, 1989.

Kurakado, M., *Nucl. Instrum. Meth. A* **196,** 275 (1982).

Li, L. et al., *Journal of Applied Physics* **90,** 3645-7 (2001).

McDonough, R.N., *Detection of Signals in Noise* (Academic, San Diego, 1995).

Pathria, R. K., *Statistical Mechanics* (Pergamon, Oxford, 1972), p. 98.

Peacock, A. et al., *Nature* **381,** 135-137 (1996).

Porter, F.S. et al., *Nucl. Instrum. Meth. A* **444,** 220-223, (2000).

Porter, F.S. et al., *Proc. SPIE* **3765,** 729 (1999).

Rando, N. et al., "The properties of niobium superconducting tunneling junctions as X-ray detectors", *Nucl. Instrum. Meth. A,* vol. 313, pp. 173-195, 1992.

Rogovin, D. and D. J. Scalapino, *Ann. of Phys.* **86,** 1 (1974).

Rothwarf, A. and B. N. Taylor, *Phys. Rev. Lett.* **19,** 27 (1967).

Schoelkopf, R.J. et al., *Science* **280,** 1238-42 (2001).

Segall, K. and D.E. Prober, *Physical Review B* **64,** 180508 (2001).

Segall, K., PhD Dissertation, Yale University (2000)

Szymkowiak, A.E., private communication.

Tinkham, M., *Introduction to Superconductivity* (McGraw-Hill, New York, 1996).

van Vliet, K. M. and J. R. Fassett, in *Fluctuation Phenomena in Solids,* edited by R. E. Burgess, (Academic, New York, 1965).

Verhoeve, P. et al., *Proc. 7th Int. Work. on Low Temp. Detectors,* p. 97-100 (1997).

Verhoeve, P., in *Proc. 9th Int. Work. on Low Temp. Detectors,* edited by F.S. Porter et al., pp. 559-564 (AIP Conference Proceedings, vol. **605**, New York, 2002).

Wilson, C. M. et al., *Nucl. Instrum. Meth. A* **444**, 449, (2000).

Wilson, C. M. et al., *IEEE Trans. Appl. Supercond.* **11**, 645 (2001a).

Wilson, C. M. et al., *Physical Review Letters* **87**, 067004 (2001b)

Williams, R. et al. 1996, Science with the *Hubble Space Telescope* II, ed. P. Benvenuti, F. D. Macchetto, & E. J. Schreier (Baltimore: STScI) 1996.

Appendix A: Device Fabrication Procedure

Frunzio-Wilson v. 2.1 (July 2002)
This fabrication procedure is largely just an elaboration of that developed by Michael Gaidis [Gaidis 1994].

I. Wafer Surface Oxidation

Use 2" diameter Siltronix Si wafer (max 9 each time):

a) Soak in acetone (ACE) with ultrasonic agitation for 2 minutes;
b) Rinse with de-ionized water (DI) for 1 minute;
c) Rinse in running DI until you get 13 MΩ-cm;
d) Soak in (4 : 1 = H_2SO_4 : H_2O_2) solution for 5 minutes (put acid into base and put the beaker into a Petri filled with water to handle this exothermic solution);
e) Rinse with DI for 2 minute;
f) Rinse in running DI until you get 13 MΩ-cm;
g) Soak in (100 : 1 = H_2O : HF) solution for 2 minutes (do not use glass beaker);
h) Rinse with DI for 2 minutes;
i) Rinse in running DI until you get 13 MΩ-cm;
j) Soak in (5 : 1 : 1 = H_2O : NH_4OH : H_2O_2) solution for 3 minutes;
k) Rinse with DI for 2 minutes;
l) Rinse in running DI until you get 13 MΩ-cm;
m) Soak in (5 : 1 : 1 = H_2O : HCl : H_2O_2) solution for 3 minutes;
n) Rinse with DI for 2 minutes;
o) Rinse in running DI until you get 13 MΩ-cm;
p) Soak in methanol (MET) to preserve native oxide growth (2 nm in 30 minutes, which is also the native oxide thickness in saturation at room temperature);
q) Blow dry wafers in N_2;
r) Oxidize wafers in the wet oxidation furnace at 1000 °C for 30 minutes to obtain ≈300 nm amorphous-SiO and 20 minutes in N_2 to anneal the surface.

II. Wafer Preparation for Tantalum Absorber Deposition

1) Clean the oxidized wafer:
   a) Soak in normal-methyl-2,pyrrolidinone (NMP) with ultrasonic agitation for 20 seconds;

    - b) Soak in ACE with ultrasonic agitation for 20 seconds;
    - c) Soak in MET with ultrasonic agitation for 20 seconds;
    - d) Rinse in running DI for 1 minute;
    - e) Blow dry wafer with N_2;
    - f) Spin dry wafer for 45 seconds;
2) Load the wafer in the Kurt .J. Lesker sputter system on the SiO-coated wafer holder, with a 3/64" thick and 2" diameter stainless steel disk on its back.
3) Prebake the wafer at 350 °C for at least 6 hours:
    - a) Check the Athena heater thermocouple is on #6 of the yellow connectors on top left of the chamber;
    - b) Make sure switch is on DLA (Dual Lamp Assembly);
    - c) Turn on all cooling water supplies;
    - d) Set Athena on standby (red light on) (Lamp SCR will blow if MAIN power is turned on when Athena is not in standby);
    - e) Switch on the MAIN power;
    - f) Use arrows to set temperature (350 °C);
    - g) Hit index button until you get to "-At- 02";
    - h) Hit tune to start heating process;
    - i) Open STA-3 MAIN shutter.

III. Titanium Sputter Cleaning

1) Check that the base pressure is about $3x10^{-7}$ Torr before deposition.
2) Sputter Ti from the 1" torus dc magnetron sputter gun:
   Note that the parameters should be argon=10 mTorr and 175.9 sccm; 125 W, 340 V, 0.39 A.
    - a) Switch off the ion gauge;
    - b) Set conductance controller to 4.0;
    - c) Close the conductance controller switch;
    - d) Close the CHAM INTERLOCK valve;
    - e) Switch on station gas AR2;
    - f) Switch on the capacitance manometer;
    - g) Switch on the flow controller (FC) main;
    - h) Switch on FC channel 1;
    - i) Turn the capacitance manometer controller to "AUTO";
    - j) Set pressure to 10 mTorr (1 turn=10 mTorr);
    - k) Set the switch on gun B (Titanium);
    - l) Switch on the 1.5 kW power supply;
    - m) Fix the setpoint at 40 W;
    - n) Switch on "OUTPUT" to start plasma (sky blue light);

o) Increase the setpoint to about 125 W;
p) Sputter for 40 minutes;
q) Decrease the setpoint to 40 W;
r) Set pressure to 0 mTorr;
s) Open the CHAM INTERLOCK valve.

IV. Tantalum Presputter

1) Check that the base pressure is about 3×10^{-7} Torr before deposition.
2) Presputter Ta from the 2" torus dc magnetron sputter gun:
   Note that the parameters should be argon=6 mTorr and 146 sccm; 300 W, 295 V, 1.04 A.
   a) Switch off the ion gauge;
   b) Set conductance controller to 1.5;
   c) Close the conductance controller switch;
   d) Close the CHAM INTERLOCK valve;
   e) Switch on station gas AR3;
   f) Switch on the capacitance manometer;
   g) Switch on the flow controller (FC) main;
   h) Switch on FC channel 1;
   i) Turn the capacitance manometer controller to "AUTO";
   j) Set pressure to 6 mTorr;
   k) Set the switch on gun F (Tantalum);
   l) Switch on the 1.5 kW power supply;
   m) Fix the setpoint at 40 W;
   n) Switch on "OUTPUT" to start plasma (sky blue light);
   o) Increase the setpoint to about 300 W;
   p) Presputter for 3 minutes;
   q) Decrease the setpoint to 40 W;
   r) Set pressure to 0 mTorr;
   s) Open the CHAM INTERLOCK valve.
   t) Increase the DLA temperature to 750 °C (it takes approximately 2 minutes).

V. Ion Beam Cleaning

1) Ion beam clean the wafer with the 2 cm Kaufman ion source:
   Note that the parameters should be argon=1.5 sccm; beam=500 V and 4.7 mA; accelerator=35 V and 0.1 mA; discharge=35 V and 0.74 A; cathode=6.6 A; neutralizer=4.0 A and 9.1 mA; (corresponding to a pressure of about 1.8×10^{-5} Torr and an Ar^+ current of about 330 $\mu A/cm^2$).

a) Check that the ion gauge is off;
b) Check that chamber interlock is open;
c) Check that the conductance control is set to 1.5;
d) Check that the conductance controller is closed;
e) Switch on station gas AR4;
f) Switch on channel 2 of the flow contoller (set at 1.5 sccm);
g) All knobs to zero on the ion beam power supply;
h) Source on (left switch; right one is for beam);
i) Discharge up until V = 55 V;
j) Turn up cathode until discharge current = 0.2 A (discharge light stops blinking at about 0.14 A; cathode current > 5.5 A);
k) Turn up beam voltage four turns clockwise;
l) Turn up accelerator voltage 1/2 turn clockwise;
m) Turn on beam (all control LED's should be on, none blinking);
n) Discharge voltage down to 35 V;
o) Set beam voltage to 500 V;
p) Set accelerator voltage to 35 V (if accelerator current is more than 25% of beam current, turn the beam off);
q) Turn up cathode current until beam current is 4.7 mA
r) Increase neutralizer current until probe current is zero (shutter must be closed. Neutralizer current should be around 4 A, discharge current around 5 mA);
s) Adjust beam current if necessary
t) Open shutter #4;
u) Clean for 2 minutes;
v) Close shutter 4 or move the sample on the next target;
w) Turn off beam;
x) Turn off knobs (in order: accelerator, beam, cathode and discharge);
y) Switch off channel 2 of the flow contoller;
z) Switch off station gas AR4.

VI. Tantalum Absorber Deposition

1) Sputter Ta to desired thickness:

Note that the parameters should be argon=6 mTorr and 145.6 sccm; 960 W, 410 V, 2.3 A; rate=3.4 nm/s. Source to sample distance is approximately 3 inches.

a) Close the CHAM INTERLOCK valve;
b) Set pressure to 6 mTorr;
c) Switch on "OUTPUT" to start plasma (sky blue light);

d) Increase the setpoint at 960 W;
e) Open shutter F (sputtering time 175 seconds for Ta thickness of 600 nm);
f) To shut down, close shutter F;
g) Ramp down power level to zero;
h) Turn off the 1.5 kW power supply;
i) Turn the capacitance manometer controller to "CLOSE";
j) Switch off the flow controller channel 1;
k) Switch off the flow controller main;
l) Switch off the capacitance manometer;
m) Switch off station gas AR3;
n) Open the CHAM INTERLOCK valve;
o) Cool sample at least 2 hours in vacuum, to <300 °C:
p) Close STA-3 MAIN shutter;
q) Set Athena on standby (red light on) (otherwise SCR will blow);
r) Switch off the MAIN power;
s) Use arrows to set temperature (350 °C);
t) Remove the sample and turn off all cooling water supplies.

VII. Tantalum Absorber Patterning by Positive Photoresist

a) Clean the tantalum surface:
    a) Soak in ACE with ultrasonic agitation for 20 seconds;
    b) Soak in MET with ultrasonic agitation for 20 seconds;
    c) Rinse in running DI for 1 minute;
    d) Blow dry wafer with N_2;
    e) Spin dry wafer for 45 seconds;
b) Spin Shipley 1813 photoresist at 4100 rpm for 45 seconds (thickness≈1.2 μm).
c) Bake on hotplate at 90 °C for 180 seconds.
d) Expose in HTG contact mask aligner for 16 seconds (4.87 mW/cm^2) with ABSORBER mask.
e) Develop in 100 ml of (1 : 1 = H_2O : Microposit Developer) for 90 seconds.
f) Soak in DI bath immediately to stop the development.
g) Rinse in running DI for 1 minute.
h) Bake on hotplate at 110 °C for 10 minutes.
i) Etch in 52 ml of (3 : 1=Kepro etch solution ($FeCl_3$, HCl) : HF) for 108 seconds, etch rate is about 5.5 nm/sec (it start to clear the Ta after 95 seconds then the endpoint is determined by inspection).
j) Soak in DI bath immediately to stop the etching.

k) Rinse in running DI for 1 minute.
l) Blow dry wafer with N_2.
m) Spin dry wafer for 45 seconds.
n) Strip the photoresist:
  a) Soak in ACE/PH with ultrasonic agitation for 20 seconds;
  b) Soak in ACE with ultrasonic agitation for 20 seconds;
  c) Soak in MET with ultrasonic agitation for 20 seconds;
  d) Rinse in running DI for 1 minute;
  e) Blow dry wafer with N_2;
  f) Spin dry wafer for 45 seconds;

VIII. Niobium Contact Patterning by Negative Photoresist

1) Clean the wafer:
  f) Soak in ACE with ultrasonic agitation for 20 seconds;
  g) Soak in MET with ultrasonic agitation for 20 seconds;
  h) Rinse in running DI for 1 minute;
  i) Blow dry wafer with N_2;
  j) Spin dry wafer for 45 seconds;
2) Spin Hoechst-Celanese AZ5218E photoresist at 3500 rpm for 45 seconds (thickness≈1.6 μm).
3) Bake on hotplate at 90 °C for 180 seconds.
4) Expose in HTG contact mask aligner for 5.5 seconds (4.87 mW/cm^2) with CONTACT mask.
5) Bake on hotplate at exactly 110 °C for 150 seconds.
6) Blanket expose the wafer in HTG contact mask aligner for 65 seconds.
7) Develop in 100 ml of (1 : 1 = H_2O : Microposit Developer) for 65 seconds.
8) Soak in DI bath immediately to stop the development.
9) Rinse in running DI for 1 minute.
10) Spin dry wafer for 45 seconds.

IX. Niobium Contact Deposition and Lift-off

1) Load sample in Leaker sputter system on a stainless wafer holder, with a 3/64" thick, 2" diameter copper disk on the back of the wafer.
2) Check that the base pressure is about 3x10^{-7} Torr before deposition.
3) Turn on all cooling water supplies.
4) Presputter Nb from the 2" torus dc magnetron sputter gun for 3 minutes (following the presputtering procedure II.4).

Note that the parameters should be argon=11 mTorr and 152.7 sccm; 290 W, 290 V, 1 A.

a) Switch off the ion gauge;
b) Set conductance controller to 4.0;
c) Close the conductance controller switch;
d) Close the CHAM INTERLOCK valve;
e) Switch on station gas AR3;
f) Switch on the capacitance manometer;
g) Switch on the flow controller (FC) main;
h) Switch on FC channel 1;
i) Turn the capacitance manometer controller to "AUTO";
j) Set pressure to 11 mTorr;
k) Set the 1 kW switch to gun E (Niobium);
l) Switch on the 1 kW power supply;
m) Fix the setpoint at 40 W;
n) Switch on "OUTPUT" to start plasma (light violet);
o) Increase the setpoint to about 330 W (Note that power will only reach ≈290 W with shutter closed);
p) Presputter for 3 minutes;
q) Decrease the setpoint to 40 W;
r) Set pressure to 0 mTorr;
s) Open the CHAM INTERLOCK valve.

5) Ion beam clean the wafer for 1.5 minutes (following the procedure V.).
6) Sputter Nb to desired thickness:

Note that the parameters should be argon=11 mTorr and 152.7 sccm; 330 W, 330 V, 1 A; rate=1.0 nm/s. Source to sample distance is approximately 3 inches. The Nb thickness should be kept below 250 nm to reduce the probability of photoresist cracking or peeling during the sputter deposition.

a) Close the CHAM INTERLOCK valve;
b) Set pressure to 11 mTorr;
c) Switch on "OUTPUT" to start plasma (light violet);
d) Increase the setpoint at 330 W;
e) Open shutter E (sputtering time 240 seconds for Nb thickness of 240 nm);
f) To shut down, close shutter E;
g) Ramp down power level to zero;
h) Turn off the 1 kW power supply;
i) Turn the capacitance manometer controller to "CLOSE";

- j) Switch off the flow controller channel 1;
- k) Switch off the flow controller main;
- l) Switch off the capacitance manometer;
- m) Switch off station gas AR3;
- n) Open the CHAM INTERLOCK valve;

7) Remove the sample and turn off all cooling water supplies.
8) Lift-off the photoresist:
   - a) Soak in ACE/PH for at least 30 minutes to gently remove the Nb-coated photoresist;
   - b) Soak in ACE with ultrasonic agitation for 20 seconds;
   - c) Soak in MET with ultrasonic agitation for 20 seconds;
   - d) Rinse in running DI for 1 minute;
   - e) Blow dry wafer with N_2;
   - f) Spin dry wafer for 45 seconds;

X. Aluminum Trilayer Deposition

1) Clean the tantalum surface:
   - a) Soak in ACE with ultrasonic agitation for 20 seconds;
   - b) Soak in MET with ultrasonic agitation for 20 seconds;
   - c) Rinse in running DI for 1 minute;
   - d) Blow dry wafer with N_2;
   - e) Spin dry wafer for 45 seconds;
2) Load sample in Lesker sputter system on a stainless wafer holder, with a 3/64" thick, 2" diameter copper disk on the back of the wafer.
3) Check that the base pressure is about 3×10^{-7} Torr before deposition.
4) Premelt aluminum evaporation filaments and tune the deposition rate during the ion beam cleaning:
   - a) Al filaments are on sources 2 to 4 (big Al switch). Lambda power supply is used for sources 1 to 3 and the big supply work for source 4;
   - b) Crystal monitor on , enter 8, FLM# 1 (density = 2.70 g/cc, z-ratio = 1.08, tooling factor = 399%, sensor = 1);
   - c) All cooling water on;
   - c) Preheat Al filaments at (20 A, 0.16 V), (40 A, 0.4 V), (60 A, 1.3 V), (80 A, 2.0 V) for 2 minutes each. Al evaporation starts with current around 80 A.
5) Ion beam cleaning of the wafer for 2 minutes (following the procedure V.).

6) Immediately, begin deposition of aluminum at a rate of about 20 nm/s for about 8 seconds to the desired trap layer thickness of 160 nm:
   a) Power supply to about 100 A;
   b) J-arm to source n (no shutter over thermal evaporator);
   c) After evaporation turn away sample and decrease current to zero.
7) Cool the ion gun and aluminum filaments for 5 minutes.
8) Isolate the vacuum chamber and sample from all pumps.
9) Bleed in pure oxygen to a pressure of = 500 mTorr (as desired) in = 1 min:
   a) Open Station gas 4;
   b) Turn on channel 3, 100 sccm, should reach 500 mTorr within 30 sec;
   c) Turn off channel 3.
10) Oxidize at nominally room temperature for ≈ 2 hours (as desired).
11) Open cryopumps directly to the chamber, pump on the chamber for 5 minutes, until the base pressure is below 2×10^{-6} Torr.
12) Warm up a new set of aluminum evaporation filaments; evaporate the counterelectrode at 2 nm/s for about 40 seconds to the desired counterelectrode layer thickness of 80 nm.
13) Immediately remove sample from vacuum chamber.

XI. Aluminum Trilayer Patterning

1) Clean the tantalum surface:
   a) Soak in ACE with ultrasonic agitation for 20 seconds;
   b) Soak in MET with ultrasonic agitation for 20 seconds;
   c) Rinse in running DI for 1 minute;
   d) Blow dry wafer with N_2;
   e) Spin dry wafer for 45 seconds;
2) Spin Shipley 1813 photoresist at 4000 rpm for 45 seconds (thickness≈1.25 μm).
3) Bake on hotplate at 90 °C for 180 seconds.
4) Expose in HTG contact mask aligner for 16 seconds (4.87 mW/cm^2) with TRAP mask.
5) Develop in 100 ml of (1 : 1 = H_2O : Microposit Developer) for 70 seconds.
6) Soak in DI bath immediately to stop the development.
7) Rinse in running DI for 1 minute.
8) Bake on hotplate at 110 °C for 10 minutes.

9) Etch in 75 ml of (8 : 4 : 1 : 1 = H_3PO_4 : CH_3COOH : HNO_3 : H_2O) (white bottle, acid cabinet) at 45 °C for 83 seconds (rate ≈ 3.7 nm/s, endpoint determined by inspection).
10) Soak in DI bath immediately to stop the etching.
11) Rinse in running DI for 1 minute.
12) Blow dry wafer with N_2.
13) Spin dry wafer for 45 seconds.
14) Test the result of this etch with the Dektak to make sure the tunnel barrier has been etched through.
15) Pattern the counterelectrode electrode with positive resist as above (**XI.1-8** with COUNTERELECTRODE mask).
16) Etch in 75 ml of (8 : 4 : 1 : 1= H_3PO_4 : CH_3COOH : HNO_3 : H_2O) (white bottle, acid cabinet) at 25 °C for 120 seconds (rate ≈ 7.6 nm/s).
17) Soak in DI bath immediately to stop the etching.
18) Rinse in running DI for 1 minute.
19) Blow dry wafer with N_2.
20) Spin dry wafer for 45 seconds.
21) Test the result of this etch with the Dektak.

XII. SiO Patterning and Deposition

1) Pattern the insulating layer with negative resist as above (**VIII.** with VIA mask).
2) Load sample in Lesker sputter system on a stainless wafer holder, with a 3/64" thick, 2" diameter copper disk on the back of the wafer. This disk should be fastened to the wafer with Dow Corning vacuum grease. The base pressure should be less than $3\text{x}10^{-7}$ Torr before the deposition.
3) Preheat the special SiO evaporation boat (R. D. Mathis model SO-10), and evaporate at 1.5 nm/s for about 167 seconds to the desired insulating layer thickness of 250 nm (>150 nm gives a pinhole-free insulating film):
   a) Use DC (Lambda) power supply and SiO boat at source 1;
   b) Crystal monitor on, enter 8, FLM# 2 (density = 2.13 g/cc, z-ratio = 0.87, tooling factor = 270%, sensor = 1);
   c) All cooling water on;
   d) Preheat SiO boat at (100 A, 0.5 V) and (200 A, 1.2 V) for 3 minutes (SiO starts to evaporate around 300 A). SiO is a getter, so the pressure will drop about one order of magnitude when crud on SiO is gone and SiO is being evaporated;

    - e) Turn DC power supply to about (330 A, 2.4 V);
    - f) Turn J-arm to source 1 (no shutter over thermal evaporator);
    - g) Crystal monitor thickness and rate are 1/2 of the actual thickness and rate.
4) Immediately remove the sample from the vacuum chamber.
5) Soak in ACE/PH for at least 30 minutes to gently remove the SiO-coated photoresist.
6) Ultrasound agitation for 20 seconds to ensure liftoff of junction via regions.
7) Clean the sample back and the copper disk with NMP to remove the vacuum grease, scrubing them if necessary.
8) Clean the sample:
    - a) Soak in ACE with ultrasonic agitation for 20 seconds;
    - b) Soak in MET with ultrasonic agitation for 20 seconds;
    - c) Rinse in running DI for 1 minute;
    - d) Blow dry wafer with N_2;
    - e) Spin dry wafer for 45 seconds;
9) Verify liftoff with microscope.

XIII. Aluminum Wiring Deposition and Patterning

1) Pattern the wiring layer with negative resist as above (**VIII.** with WIRINGX mask).
2) Deposit the wiring layer as above at a rate of 3 nm/s (**X. 2-7**) for about 100 seconds to the desired wiring layer thickness of 300 nm.
3) Immediately remove the sample from the vacuum chamber.
4) Soak in ACE/PH for at least 30 minutes to gently remove the Al-coated photoresist.
5) Ultrasound agitation for 20 seconds to ensure liftoff.
6) Clean the sample back and the copper disk with NMP to remove the vacuum grease, scrubing them if necessary.
7) Clean the sample:
    - a) Soak in ACE with ultrasonic agitation for 20 seconds;
    - b) Soak in MET with ultrasonic agitation for 20 seconds;
    - c) Rinse in running DI for 1 minute;
    - d) Blow dry wafer with N_2;
    - e) Spin dry wafer for 45 seconds;
8) Verify liftoff with microscope.

Appendix B: Important Runs and Device Parameters

Important Runs

Date	Device	Significance
6/23/99	OPS-F99-6B	First detection of single optical/UV photons
12/14/99	OPS-F99-6B	Achieve R~12 at 5 eV
1/5/00	OPS-F99-6B	Achieve R~12 at 5 eV for second time
10/12/00	OPS-E00-6B	Measurement of thermodynamic fluctuations
8/28/01	OPS-E00-7A	First meas. of non-back. device with laser
3/5/02	OPS-E00-10A	Detailed meas. of non-back. resolution

Wafer Parameters

Layer	F99 Thickness (nm)	E00 Thickness (nm)
Silicon Wafer	305 μm	510 μm
Si0$_2$ Buffer	600	300
Absorber (Ta)	570	545
Contact (Nb)	240	225
Trap	160	200
CounterElec.	80	80
Insulator	270	220
Wiring	365	275

Important Devices

Device	Sing./Doub.	Absorber (μm^2)	A$_J$ (μm^2)	Back/Non
OPS-F99-6B	Double	10x100	100	Back
OPS-E00-6B	Double	10x100	100	Back
OPS-E00-7A	Single	10x100	100	Non
OPS-E00-10A	Single	20x50	100	Converted