Near-Infrared Instrumentation and Millimeter-Wave Simulations For Cosmological Surveys

by

Tomasz P. Biesiadzinski

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Physics)
in the University of Michigan
2013

Doctoral Committee:

- Professor Gregory Tarlé, Chair
- Associate Professor Dragan Huterer
- Professor Wolfgang Lorenzon
- Assistant Professor Jeff McMahon
- Assistant Professor Christopher John Miller

Poets say science takes away from the beauty of the stars – mere globs of gas atoms. I, too, can see the stars on a desert night, and feel them. But do I see less or more?

–Richard P. Feynman

A picture of Khalid's hand, the first image taken with the Spots-O-Matic.

I'm glad I did it, partly because it was worth it, but mostly because I shall never have to do it again

–Mark Twain

ACKNOWLEDGMENTS

It took the support and advice of many people to allow me to reach the milestones of finishing my studies in graduate school as well as completing this dissertation.

First and foremost I would like to thank my advisor Greg Tarlé. He has continued to provide me with guidance, encouragement and help whenever I needed them while simultaneously giving me the freedom to explore new concepts, try out new techniques and, occasionally, fail. It is thanks to Greg and his optimism sustained through some disappointments that I leave Michigan with the confidence in my future as a physicist. I would also like to thank Wolfgang Lorenzon for his guidance with my first real research project, willingness to argue with a student who frankly didn't know that much at the time and his good natured candor. Michael Schubnell has shown an enormous amount of patience with me and my often inane questions while always making me feel like I contribute something. This was of course very much appreciated. Thanks to Jeff McMahon for his advice, confidence in my abilities and enthusiasm about our work together. I had the pleasure of working with Chris Miller who was always full of friendly advice and who helped me a great deal in navigating the politics of my chosen profession. Thanks to Dragan Huterer for helping out a budding experimentalist (sort of) understand likelihoods and Fisher matrices and for the occasional chats.

I want to express my gratitude to Curtis Weaverdyck who helped design, build and operate various experiments in the detector lab that made this dissertation possible.

Thanks also to Brian Nord, Chris Bebek, Roger Smith and Laurie Shaw for their collaboration and insights during my studies. I would also like to thank the United States taxpayers who funded much of my work through the U.S. Department of Energy and I hope that I was able to and will continue to provide something in return.

I want to thank the undergraduate students who have worked with me over the years. I appreciate all of the help provided by Robert Newman when I was first starting to work in the lab as well as his insistence on dragging me outside to have a beer once in a while. Big thanks to Michael Howe, Khalid Jawed and Joshua Larson who continued laboring on the Spots-O-Matic; it would not have been possible without them.

Finally, I would like to thank all the friends that I managed to make despite my best efforts; they helped to make the bad times bearable and the good times memorable. And a big thanks to my family for always believing in me.

TABLE OF CONTENTS

Chapter

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

2dFGRS Two-degree-Field Galaxy Redshift Survey

400SD 400 Square Degree ROSAT PSPC galaxy cluster survey

ACT *Atacama Cosmology Telescope*

ADDGALS Adding Density Determined GAlaxies to Lightcone Simulations

ADU analog to digital units

AGN Active Galactic Nuclei

ASICs application-specific integrated circuits

BAO Baryon Acoustic Oscillation

BBN Big Bang Nucleosynthesis

BCC Blind Cosmology Challenge

BOSS Baryon Oscillation Spectroscopic Survey

CAMB Code for Anisotropies in the Microwave Background

CDM Cold Dark Matter

CDS Correlated Double Sampling

CfA Harvard–Smithsonian Center for Astrophysics

CMB Cosmic Microwave Background

COBE *COsmic Background Explorer*

CTIO Cerro Tololo Inter-American Observatory

DECam Dark Energy Camera

DES Dark Energy Survey

DESI Mid-Scale Dark Energy Spectroscopic Instrument

eBOSS Extended Baryon Oscillation Spectroscopic Survey

ELGs Emission Line Galaxies

ESA European Space Agency

ESO European Southern Observatory

FLRW Friedman–Lemaître–Robertson–Walker

FPA focal plane array

FWHM full width at half maximum

GNFW generalized NFW

GR General Relativity

HaRMLSS HAlo–Resolved Millimeter-wave Layered Sky Simulation

HAWAII HgCdTe Astronomy Wide Area Infrared Imager

H2RG HAWAII-2RG

HgCdTe Mercury-Cadmium-Telluride

HEALPix Hierarchical Equal Area isoLatitude Pixelization

HSE hydrostatic equilibrium

HST *Hubble Space Telescope*

ICM intracluster medium

InGaAs indium gallium arsenide

IR infrared

JDEM *Joint Dark Energy Mission*

kSZ kinetic Sunyaev–Zel'dovich

LSS Large-Scale Structure

LRG Luminous Red Galaxy

LSS Large Scale Structure

LSST *Large Synoptic Survey Telescope*

Lyα Lyman-α

MF Matched Filter

MCXC Meta-Catalog of X-ray Detected Clusters of Galaxies

ND neutral density

NFW Navarro–Frenk–White

NICMOS Near Infra-Red Camera and Multi-Object Spectrometer

NIR near-infrared

NORAS Northern ROSAT All-Sky

PDF probability distribution function

PRF pixels response function

PSD power spectral density

PSF point spread function

QE quantum efficiency

QTH Quartz-Tungsten-Halogen

RASS ROSAT All-Sky Survey

RSD redshift space distortion

SDSS Sloan Digital Sky Survey

Si Silicon

S/N signal to noise

SNAP *Supernovae Acceleration Probe*

SNe Supernovae

SPT *South Pole Telescope*

SSS Small Scale Structure

STIS Space Telescope Imaging Spectrograph

SUR Sample-Up-the-Ramp

SZ Sunyaev–Zel'dovich

SZA Sunyaev–Zel'dovich Array

TES transition edge sensor

TIS Teledyne Imaging Sensors

tSZ thermal Sunyaev–Zel'dovich

VISTA *Visible and Infrared Survey Telescope for Astronomy*

VHS VISTA Hemisphere Survey

WFC3 Wide Field Camera 3

WFIRST *Wide Field Infrared Survey Telescope*

WMAP *Wilkinson Microwave Anisotropy Probe*

ABSTRACT

Near-Infrared Instrumentation and Millimeter-Wave Simulations For Cosmological Surveys

by

Tomasz P. Biesiadzinski

Chair: Professor Gregory Tarlé

The evolution of the Universe is well characterized by the concordance ΛCDM cosmological model where structure formation is seeded by cold dark matter and accelerated expansion is driven by the cosmological constant. Understanding the history and fate of the Universe requires precise measurements of cosmological parameters. Finding them inconsistent may lead to a more fundamental physical theory. I explore observable probes of cosmological parameters as well as instrumental effects that may obfuscate them.

I develop a framework for simulating millimeter-wave skies including galaxy clusters' Sunyaev–Zel'dovich (SZ) signature. This framework includes astrophysical and instrumental effects. Its primary use is in testing systematic effects resulting from joining intrinsic profile variations and mass dependencies with observational uncertainties and signal extraction techniques as well as multi-wavelength studies. I demonstrate that the signal recovered using Matched Filter is very sensitive to SZ profile shapes and potentially leads to biases.

I then consider the impact of galaxy cluster selection and characterization in the maxBCG cluster catalog on recovering a stacked SZ signal in light of recently measured biases. I find that accounting for the mass calibration uncertainty and mis-centering of galaxy clusters may explain the majority of the observed discrepancy. In addition, contrary to others' findings, I conclude that the X-ray sub-sample of maxBCG clusters is similarly affected. My findings suggest that the SZ signal can indeed serve as an alternate mass calibration technique.

I finally focus on instrumental effects in near-infrared (NIR) detectors designed for large surveys of the cosmos. I first characterize the flux dependent non-linearity known as reciprocity failure and find that it can be as large as 10% per decade in flux change but is suppressed by cooling the detectors. I then thoroughly study the quantum efficiency (QE) of a single NIR device under different environmental and illumination conditions and conclude that it can vary significantly. Careful accounting of various sources of uncertainty suggests that some observers may be too confident in the quality of their QE measurements.

CHAPTER 1

Introduction to Modern Cosmology

The modern view of the universe that emerged over the past century has allowed us to understand its various, seemingly disconnected characteristics. The past two decades in particular have helped cement our current standard model of cosmology, the concordance ΛCDM model. Here Λ represents the currently dominant dark energy accelerating the expansion of the universe and CDM represents cold dark matter that drove structure formation in the earlier epochs of the universe and lead to the creation of galaxies, stars and eventually ourselves.

1.1 Overview of the Dissertation

In this introduction I first provide a brief overview of historical observations that allowed us to develop the concordance cosmological model (§1.2) and the timeline of the evolution of the universe (§1.3). I then discuss the theoretical framework of the ΛCDM model along with its observational underpinnings (§1.4) and the major probes of dark energy (§1.5). Next I provide a brief overview of galaxy clusters, their formation, observables and use for cosmological studies (§1.6) to motivate my work on Sunyaev–Zel'dovich (SZ) cluster signatures. Following that I summarize some of the major optical (including near-infrared (NIR)) and SZ surveys of the universe (§1.7). I conclude the introduction with a discussion of NIR detectors utilized by some of those surveys and provide examples of detector effects that motivated us to investigate them in detail for space telescope applications (§1.8).

The remainder of this dissertation covers original research I conducted. It can be roughly split into studies involving the SZ effect (Chapters 2 and 3) and NIR detector characterization (Chapters 5 and 6). Combined, this work has allowed me to learn about all levels of the process of investigation into dark energy, from raw data acquisition to cosmological interpretation of observed astronomical sources and structures.

Chapter 2 describes the construction of the HAlo–Resolved Millimeter-wave Layered Sky Simulation (HaRMLSS) framework. Essentially, HaRMLSS paints SZ profiles on dark matter halo positions obtained from N-body simulations and dresses them with backgrounds, foregrounds and instrumental noise. It is very flexible in the signal and noise implementations. Mostly HaRMLSS allows for detailed systematic studies that include the interplay between theory and observational techniques. I conclude this chapter with applications of this framework to study biases caused by SZ profile shapes and scaling relations. This work was performed in collaboration with Jeff McMahon and Christopher Miller as well as with the input of the DES and SPT collaborations.

The HaRMLSS framework is utilized in Chapter 3 in order to study the impact of optical cluster-finder selection and characterization on stacked SZ signals. Such stacking may provide an additional means of mass calibration of optically detected clusters. An analysis performed by members of the *Planck* collaboration where the *Planck* measured signal was stacked on optical cluster positions found a significant signal deficit (Planck Collaboration et al. 2011d). I simulated optical cluster catalog systematics in selecting signal from simulated SZ maps in order to study this effect. This chapter reproduces and expands on Biesiadzinski et al. (2012). This work was performed in collaboration with Jeff McMahon, Christopher Miller, Brian Nord and Laurie Shaw.

Chapter 4 discusses the future upgrades of the HaRMLSS framework as well as the theoretical and observational difficulties in multi-wavelength study of clusters of galaxies. These must be addressed in order to fully utilize this powerful probe of cosmology.

In Chapter 5 I present a detailed study of a flux–dependent non-linearity known as reciprocity failure observed in HgCdTe NIR detectors. This non-linearity can have profound effects on Supernovae (SNe) cosmology as well as on weak lensing reconstruction and must be understood if these devices are to be used to extend precision cosmological measurements to high redshifts. This chapter condenses and expands on Biesiadzinski et al. (2011b,a). This work was performed in collaboration with Wolfgang Lorenzon, Robert Newman, Michael Schubnell, Gregory Tarlé and Curtis Weaverdyck.

In Chapter 6 I discuss a detailed characterization of the quantum efficiency of a NIR device, H2RG-236. This study illustrates that quantum efficiency is a complicated function of a detector's operating conditions such as temperature and bias voltages as well as a function of the luminosities of the sources the detector is observing. This work has been submitted for publication in Biesiadzinski et al. (2013) and was performed in collaboration with Wolfgang Lorenzon, Michael Schubnell, Gregory Tarlé and Curtis Weaverdyck.

Chapter 7 summarizes the detector effects presented in this dissertation. It also discusses their potential impact on future measurements and how to mitigate it.

Some readers may also be interested in the content of the attached appendices. In Appendix A I describe work performed in optimizing a Matched Filter for use with SZ stacking on optical cluster centers. Appendix B discusses various statistical tools developed for characterizing detector defects. In Appendix C I discuss the first Spots-O-Matic measurement of a detector and its initial results.

1.2 A History of Looking Outward

Humanity has always been fascinated by the nighttime sky. It helped us keep time, plant crops and find our way. It also continuous to fill us with awe. But it was only relatively recently that we have been truly able to explore its mysteries.

The old Earth-centric celestial spheres were first questioned in modern history by Copernicus who suggested that the Sun, and not the Earth lies at the center of the universe (Brown 2007; Stanek 2009). Galileo looked out through the telescope and saw mountains on the moon, rings around Saturn and moons orbiting Jupiter further destroying the idea of a "perfect" universe with the Earth at its center. Kepler, starting with the techniques and data of Brahe, discovered the rules of planetary motion and the elliptical nature of orbits, and Newton figured out how celestial bodies interact via gravity to make this apparent motion possible. Observers continued to discover more about our solar system and beyond. At the same time, the development of the general laws of physics explained how our universe could be the way we see it via the theory of *General Relativity* (e.g. Einstein 1915, 1917) and experimental discoveries such as Hubble's discovery of the expanding universe (Hubble 1929). Messier and Herschel in the eighteenth century noticed concentrations of "nebulae" in the sky that are now understood as galaxy clusters (Biviano 2000). Once the extragalactic origin of galaxies became known (e.g. Crommelin 1918; Hubble 1925a,b) galaxy clusters became intensely studied. The first clues pointing to the existence of *dark matter* emerged from velocity dispersions of clusters of galaxies (e.g. Zwicky 1937) that were difficult to explain with baryonic matter alone. Eventually the first of many detailed cluster catalogs were compiled (Abell 1958) to allow for their systematic study. Additional observations of galaxy rotation curves (e.g. Rubin & Ford 1970) provided further proof of the existence of dark matter. The discovery of the Cosmic Microwave Background (CMB) (Penzias & Wilson 1965; Dicke et al. 1965) in 1965 strengthened the idea of the Big Bang theory. Further measurements of the CMB (e.g. Smoot et al. 1992; Bennett et al. 2003; Spergel et al. 2003; Hinshaw et al. 2012; Keisler et al. 2011; Das et al. 2011; Planck Collaboration et al. 2013a,b, and many more) became the foundations of modern cosmology. In addition, while CMB measurements were pointing towards a flat universe, galaxy cluster observations were indicating that the matter energy density was far from sufficient to prevent an open universe (e.g. Maddox et al. 1990; Efstathiou et al. 1990; White et al. 1993). The case was finally settled (while stirring up a whole other set of issues) with the confirmation of the accelerating expansion of the universe with Type Ia SNe data (Perlmutter et al. 1997, 1999; Riess et al. 1998). And now, our story continues...

In many ways our deeper understanding of the universe has been a humbling experience. Earth, and therefore humanity, has been moved away from the center of the universe to an orbit around the Sun, the Solar system has been moved to the outskirts of the Milky Way and the Milky Way became just one of many galaxies, insignificant in the grand scheme of things. To me though, this means that we are not just some isolated beings looking up. Instead, we are all part of this universe. And that is pretty cool.

1.3 The Universe So Far

It is now time to look at the picture of the universe that the observations discussed above allowed us to construct. The true beginning of the universe remains unknown as our observations have so far failed to witness it and our physical models remain speculative at best. However it only takes a tiny fraction of a second before our theory catches up to reality and we can make testable predictions that have so far withstood all challenges. In another 400,000 years, a mere drop in the bucket compared to the 13.8 billion year age of the universe, we could finally see the light that would reveal to us so much about its history. The timeline of the evolution of the universe is pictured in Figure 1.1.

Shortly after its launch, the universe was extremely hot and dense, largely described by physics beyond the Standard Model of particle physics. But it did not remain that way for long (Nord 2010). It is thought that the universe underwent a short period of exponential growth known as *Inflation* (Guth 1981) driven by a

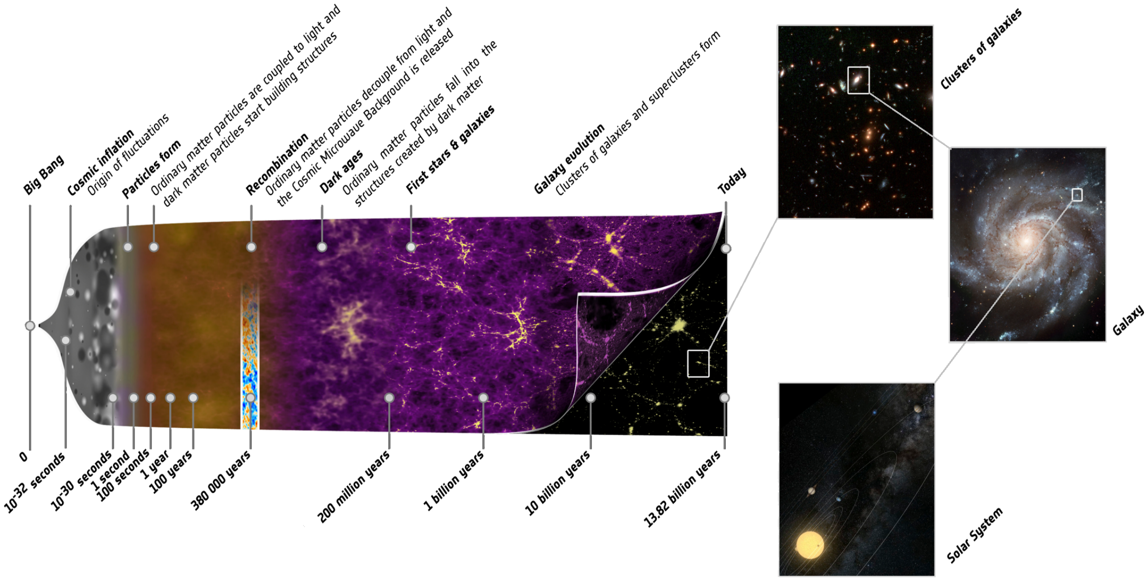


Figure 1.1: The timeline of the universe according to the current concordance cosmological model. Inflation imprints quantum fluctuations that arose during the big bang onto the primordial CMB. These fluctuations grow in time into stars, galaxies and clusters. Space itself expands at a decreasing rate until dark energy density becomes dominant and drives its accelerated expansion. Image credit: European Space Agency & C. Carreau (ESA & Carreau 2013).

theorized particle called the *inflaton* which does not appear in the Standard Model. Inflation is credited with being responsible for the homogeneity and isotropy of the universe on even the largest scales (see §1.4.1). The decay of the inflaton in a process called *reheating* is thought to give rise to the familiar content of the universe; photons, quarks, gluons, leptons and dark matter, an as yet unidentified form of matter that does not appear to interact electromagnetically. Initially, the energy density of relativistic species, photons and neutrinos, constituted a large fraction of the total energy density of the universe and hence decelerated its expansion. Shortly thereafter matter, especially dark matter, became the dominant species. As the universe continued to expand and cool, quarks and gluons formed protons and neutrons. Then, after about ten minutes the process of *Big Bang Nucleosynthesis (BBN)* (Steigman 2007) synthesized light nuclei like Deuterium, Helium and Lithium. Once the universe cooled sufficiently, at about a cosmological redshift of 1100 and at the age of 380,000 years the nuclei were able to capture free electrons to form atoms in the process known as *recombination*. At that point in space-time, known as the *surface of last scattering*, the universe became transparent to light. Photons from that era have been streaming through the universe ever since with little interaction with matter allowing us to image the primary CMB. The universe continued to expand in the post-Inflation era though the rate of this expansion was decreasing under the pull of gravity until the mysterious dark energy became sufficient to overcome it several billion years ago. Current measurements are consistent with a cosmological constant, equivalent to a constant vacuum energy density, as the explanation of dark energy though more complicated theories exist (see §1.5). Soon after *recombination*, baryonic matter fell into dark matter potential wells seeded by quantum fluctuations prior to inflation (e.g. Guth & Pi 1982) and formed the first stars and galaxies ending the dark ages of the universe about 400 million years after its launch. In time larger structures formed culminating in clusters of galaxies inhabiting massive dark matter halos. Meanwhile, light from the first stars *reionized* the atoms. Successive generations of stars forming and dying in SNe explosions enriched the interstellar medium with heavy elements eventually leading to the formation of rocky planets and finally, ourselves.

1.4 Modern Cosmological Theory

Cosmology is, to a large degree, the study of the history and future evolution of space-time itself. It is usually assumed to be described by the theory of General Relativity (GR) developed by Einstein and others (Einstein 1915) though other possibilities exist (e.g. Dvali et al. 2000; Huterer & Linder 2007). GR relates the mass and energy content of the universe with the curvature of space-time and is governed by Einstein's field equation

$$R_{\mu\nu} - \frac{1}{2}\,R\,g_{\mu\nu} - \Lambda\,g_{\mu\nu} = -\frac{8\pi G}{c^4}\,T_{\mu\nu}\ , \tag{1.1}$$

where $R_{\mu\nu}$ and R are the Ricci tensor and scalar, respectively, and $T_{\mu\nu}$ is the stress–energy tensor. The Ricci scalar is itself a function of $T_{\mu\nu}$ and the Ricci tensor is a function of the metric. Together, they act on the space-time metric $g_{\mu\nu}$ causing it to curve in the presence of matter and energy. Light as well as matter then travel along the curved paths embedded in the metric. G is the gravitational constant and c the speed of light. Finally, Λ is the cosmological constant which can act to halt the contraction of a static universe or even cause its expansion to accelerate, the purpose it serves in the current concordance ΛCDM cosmology solution to Equation 1.1. Ultimately, we wish to study the evolution of the universe by observing the effect that matter, energy and the cosmological constant have on the metric $g_{\mu\nu}$. This requires certain assumptions, specifically the homogeneity and isotropy of the universe on large scales.

1.4.1 Homogeneity and Isotropy

Initially the assumptions of homogeneity of the universe (on average, it is the same everywhere) and its isotropy (on average, it looks the same in every direction) were needed to simplify mathematical models of its evolution. Observations have since shown that these assumptions largely true but wrong in detail.

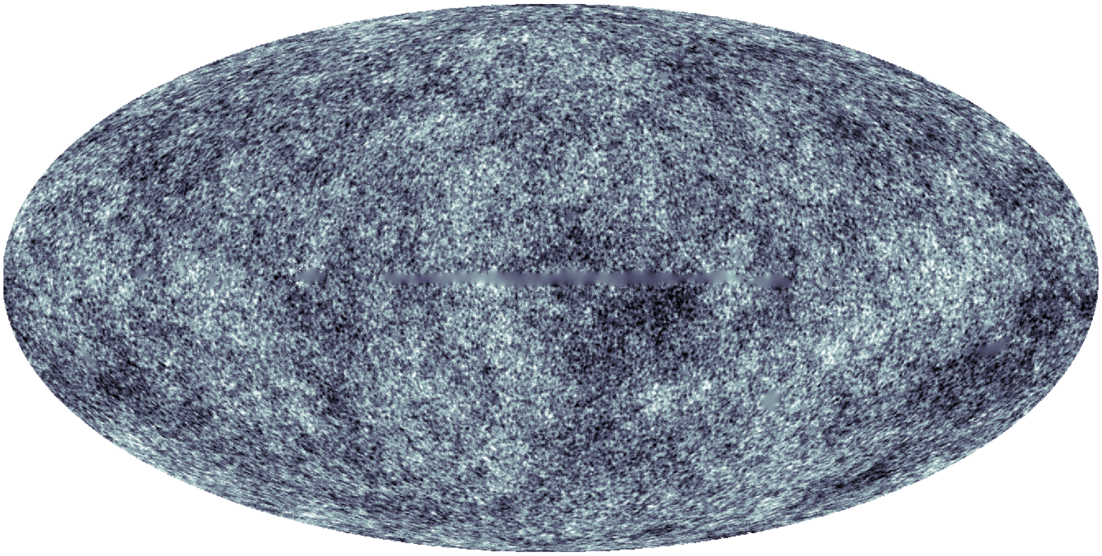


Figure 1.2: Map of the CMB at $5'$ resolution measured by the *Planck* satellite. The color scale stretches from -300 μK to 300 μK. There is a defect visible in map center due to imperfect galactic foreground removal. Data from **http://pla.esac.esa.int/pla/aio/planckProducts.html** courtesy of ESA.

The CMB is the primary evidence of large scale isotropy of the universe. It is one of the foundations of modern cosmology due to the constraints it places on so many parameters and its favoring of inflationary theory. Since its discovery in 1965 (Penzias & Wilson 1965; Dicke et al. 1965) it has been measured with greater and greater precision, most recently by the *Planck* satellite (e.g. Planck Collaboration et al. 2013a,b). It must be noted that the CMB itself is NOT isotropic; its incredible usefulness in cosmology is due to small temperature anisotropies that carry information about the flatness of the universe, matter content, baryon fraction and more. And of course, it is these anisotropies that form the seeds of structure in the universe and therefore lead to the formation of clusters, galaxies, stars, planets and ourselves. Figure 1.2 shows these anisotropies as measured by *Planck*. The RMS residual of these anisotropies from the average temperature of the CMB at a resolution of $5'$ is about 100 μK. These anisotropies are caused by the interplay of the attractive force

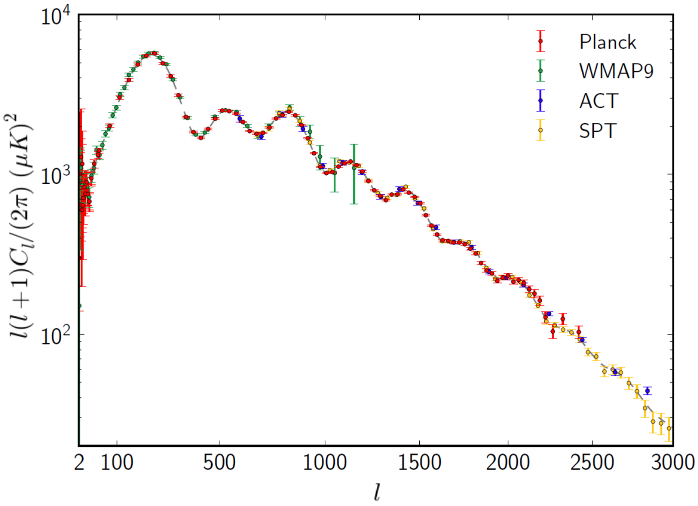


Figure 1.3: Latest measured CMB power spectrum and best fit ΛCDM model. Low ℓ data compiled from *Planck* (Planck Collaboration et al. 2013a) and *WMAP* (Hinshaw et al. 2012). High ℓ data from *SPT* (Keisler et al. 2011) and *ACT* (Sievers et al. 2013). Figure from Planck Collaboration et al. (2013a).

of gravity due to dark matter coalescing into structures and the outward pressure of the baryon–photon fluid. The power spectrum of these fluctuations can be calculated using inputs from inflationary theory (Guth & Pi 1982), GR and plasma physics. Fitting these calculated models to data can constrain cosmological parameters. The measurements of this power spectrum along with a best fit model are shown in Figure 1.3. This radial power spectrum assumes that the universe is approximately Gaussian (Guth & Pi 1982; Kolb & Turner 1990), that is, the distribution of fluctuations of the CMB on any angular scale computed over the entire CMB is Gaussian. The fact that the model matches data well is a strong confirmation of this assumption.

At large scale the CMB looks quite different. The CMB is very nearly isotropic since the RMS residual of its temperature fluctuations is only 4 parts in 10^5. The overall intensity spectrum of the CMB is shown in Figure 1.4 where it matches a thermal black body spectrum within the extremely small errors. This isotropy is strong evidence for inflation (Guth 1981) since it means that distant parts of the universe must have been in thermal contact at some point in their history.

Further evidence of isotropy can be found in the matter distribution. Reid et al. (2010) found that Luminous Red Galaxy (LRG) pairs, essentially pairs of red and very bright galaxies associated with galaxy clusters, are distributed isotropically along lines-of-sight used in their study. Of course, this is also only true after averaging over sufficiently large scales. Figure 1.5 shows the cosmic web of observed matter distribution from parts of the Two-degree-Field Galaxy Redshift Survey (2dFGRS) (Colless et al. 2001), the second Harvard–Smithsonian Center for Astrophysics (CfA) spectroscopy survey (Geller & Huchra 1989) and the Sloan Digital Sky Survey (SDSS)

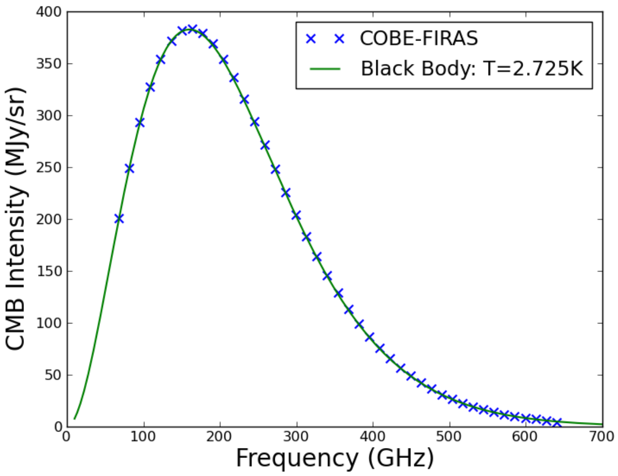


Figure 1.4: Observed CMB temperature and black body model at 2.275 K. CMB intensity from the *COBE* FIRAS spectrometer (Fixsen et al. 1996).

spectroscopy sample (Gott et al. 2005) in the local universe. Also shown is the Millennium simulation (Springel & et al. 2005) of the matter distribution. The simulation does a remarkable job of reproducing the observed features which are certainly not homogeneous on these scales. However, averaging over larger scales does in fact show that the local universe is homogeneous. Figure 1.6 shows the log_{10} of the ratio of LRG density found within a sphere of radius R relative to the density drawn from a random, homogeneous sample using SDSS data at the approximate redshift of 0.3 (Hogg et al. 2005). At small radii, there is a clear overdensity of LRGs due to the presence of galaxy clusters sourcing local inhomogeneities. As the region of space being averaged over increases, the density ratio decreases. For regions with co-moving radius larger than approximately $70\,\mathrm{h}^{-1}$Mpc this ratio becomes unity because the galaxy samples is homogeneous.

The homogeneity and isotropy of the universe has been tested quite thoroughly and until now no major deviations have been discovered. This implies that our simple treatment of Einstein's equations that follows is approximately correct. It should be noted however that recent measurements by *Planck* do seem to indicate the presence of large scale anisotropies in the CMB (Planck Collaboration et al. 2013c) at the edge of expectations based on the *cosmic variance*, the uncertainty due to the small number of large scale modes "realized" in the universe. These have been previously noticed in the *Wilkinson Microwave Anisotropy Probe* (*WMAP*) data. These anisotropies may shed light on the nature of inflation or they may be explained away as some yet unknown systematic errors. Only time, more work and more data will tell.

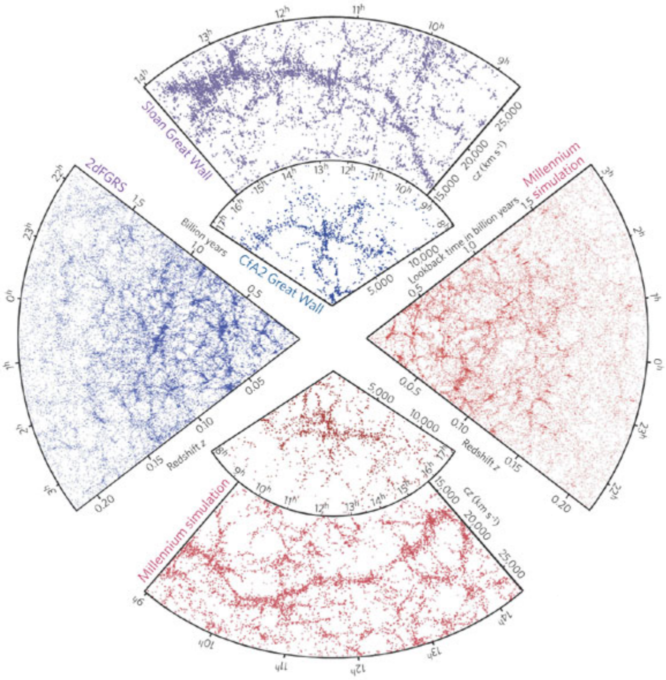


Figure 1.5: Comparison of the observed, inhomogeneous cosmic web with simulations. Left wedge: 2dFGRS extending to z of 0.22. Upper top wedge: SDSS great wall extending to z of 0.1. Lower top wedge: CfA2 great wall extending to z of 0.05. Right wedge: Millennium simulation extending to z of 0.22 matching 2dFGRS depth. Bottom wedges: Millennium simulation extending to z of 0.1 and z of 0.05 matching the depth of SDSS and CfA2, respectively. Figure from Springel et al. (2006). Note that the Millennium simulation redshift axis has been corrected.

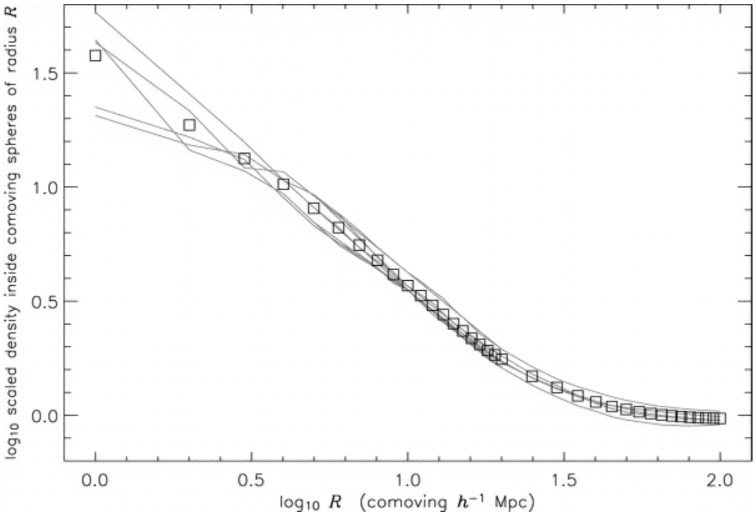


Figure 1.6: Large scale homogeneity of the matter in the local universe. Squares show the overdensities centered on 3658 LRGs as a function of averaging radius. The lines show the same for 5 samples split based on RA cuts. Figure from Hogg et al. (2005).

1.4.2 The Evolution of the Universe

Under assumptions of homogeneity and isotropy only one space-time metric $g_{\mu\nu}$ is possible (Robertson 1935; Walker 1937), the Friedman–Lemaître–Robertson–Walker (FLRW) metric (Friedman 1922; Friedmann 1924; Lemaître 1931) with a space-time interval given by

$$ds^2 = -c^2dt^2 + a(t)^2 \left(\frac{dr^2}{1-kr^2} + r^2 \left(d\theta^2 + \sin^2\theta \, d\phi^2 \right) \right) , \tag{1.2}$$

where $a(t)$ is the time dependent scale factor of the universe that traces the expansion history of the space-time. It is normalized to unity at the current time. The space coordinates (r, θ and ϕ) are co-moving (fixed in their local space). k parametrizes the curvature of the universe; for k=+1, the universe is spherical, or closed, for k=-1, it is hyper-spherical, or open and for k=0, it is Euclidean, or flat. Similarly, the assumptions of homogeneity and isotropy also allow us to approximate the content of the universe as a uniform, perfect fluid which then specifies

$$T_{\mu\nu} = \begin{bmatrix} \rho\, c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{bmatrix} , \tag{1.3}$$

where ρ is the matter or energy density and p is the pressure exerted by the perfect fluid. Using Equations 1.2 and 1.3 to solve Equation 1.1 we can arrive at the two Friedman equations

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda}{3}c^2 , \tag{1.4}$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \left(\frac{\Lambda}{3} - \frac{k}{a^2}\right)c^2 . \tag{1.5}$$

The ratio of the rate of expansion of space-time to its size $\frac{\dot{a}}{a}$ is called the Hubble parameter H. The relation between the pressure and density of a fluid is characterized by its equation of state:

$$p = w\, \rho\, c^2 , \tag{1.6}$$

with w the equation of state parameter. For cold, non-relativistic matter w=0 and for radiation and relativistic matter w=$\frac{1}{3}$.

It is helpful to separate the energy density ρ into its constituent components; cold matter (ρ_M), relativistic matter and electromagnetic radiation (ρ_R). In addition to this, we generalize the dark energy density ρ_{DE}. Instead of assuming a cosmological constant Λ with an energy density of ρ_Λ as the driving principle behind cosmic acceleration, a possibly time dependent dark energy density that causes this acceleration can be added into ρ to form total density ρ_T such that

$$\rho_T = \rho_M + \rho_R + \rho_{DE} \ , \tag{1.7}$$

with

$$\rho_{DE} = \widetilde{\rho_{DE}} + \rho_\Lambda \ . \tag{1.8}$$

$\widetilde{\rho_{DE}}$, the possible additional form of dark energy, is 0 in concordance ΛCDM universe. However other models may predict its presence in addition to or even instead of the cosmological constant.

We can now compute both the cosmological constant effective energy density ρ_Λ and pressure p_Λ. Using some algebra, equation 1.4 can be re-written as

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho - 2\frac{\Lambda}{8\,\pi\,G}c^2 + \frac{3p}{c^2}\right) \ , \tag{1.9}$$

which transforms into

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\left[\rho + \frac{\Lambda}{8\,\pi\,G}c^2\right] + \frac{3}{c^2}\left[p - \frac{\Lambda}{8\,\pi\,G}c^4\right]\right) \ . \tag{1.10}$$

This equation is very illustrative in that it allows us to define the cosmological constant energy density ρ_Λ with

$$\rho_\Lambda = \frac{\Lambda}{8\,\pi\,G}c^2 \ , \tag{1.11}$$

and the cosmological constant pressure

$$p_\Lambda = -\frac{\Lambda}{8\,\pi\,G}c^4 \ . \tag{1.12}$$

This implies that the cosmological constant energy density and pressure are related as

$$p_\Lambda = -\rho_\Lambda\, c^2 \ , \tag{1.13}$$

and therefore the cosmological constant equation of state w_Λ relating it effective energy density and pressure is

$$w_\Lambda = \frac{p_\Lambda}{\rho_\Lambda\, c^2} = -1 \ . \tag{1.14}$$

This brings up an interesting feature of the pressure due to the cosmological constant (and of generalized dark energy as well). This pressure is *negative*. It is informative to discuss exactly what this means for the expansion of the universe. First, let us discuss the more familiar components of the universe. For matter energy density ρ_M with an equation of state $w = 0$, only the gravitational potential it sources acts on the space-time causing its deceleration. Radiation density ρ_R actually acts more strongly to decelerate the universe. Not only does it source a gravitational potential but it also exerts *positive* pressure acting to slow down the expansion of the universe. This is counter-intuitive in that we tend to think of positive pressure pushing on the walls of a container like a balloon *driving* its expansion. But the opposite happens with space-time. The *negative* pressure due to the cosmological constant then acts to accelerate the expansion of the universe.

Using Equations 1.10, 1.11 and 1.12 the first Friedman Equation (1.4) becomes

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho_T + \frac{3\,p_T}{c^2}\right) \ , \tag{1.15}$$

and the second Friedman Equation (1.5) becomes

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho_T - \frac{k}{a^2}c^2 \ , \tag{1.16}$$

with Λ now implicitly included in the definitions of the total density ρ_T and total pressure p_T.

Differentiating Equation 1.16 with respect to time and substituting it into Equation 1.15 we can derive

$$\dot{\rho_T} = -3\,\frac{\dot{a}}{a}\left(\frac{p_T}{c^2} + \rho_T\right) \ , \tag{1.17}$$

which is also the continuity equation emerging from the first law of thermodynamics obtained by differentiating

$$\rho_T = \frac{E_T}{a^3} \ , \tag{1.18}$$

with respect to time in an adiabatic and number density conserving fashion. Here E_T is the fluid mass/energy and a^3 is the volume element. This means that the density is changing as the volume expands and work is being done by the pressure

$$\dot{\rho_T} = \frac{1}{a^3}\frac{dE_T}{dt} - E_T\frac{1}{a^6}\frac{d}{dt}(a^3) \ . \tag{1.19}$$

$\frac{dE_T}{dt}$ is the work done by pressure and equals to $-\frac{p_T}{c^2}\frac{d}{dt}(a^3)$. Equation 1.19 then simplifies to

$$\dot{\rho_T} = -\frac{1}{a^3}\frac{d}{dt}(a^3)\left(\frac{p_T}{c^2} + \frac{E_T}{a^3}\right) \ , \tag{1.20}$$

and finally to Equation 1.17 after performing the substitution of Equation 1.18. The fact that this process is adiabatic and number-density conserving suggests that individual components of the fluid (matter, radiation, dark energy) satisfy Equation 1.17 independently. This means that it can be solved for each energy density component without any cross-terms. This is not always the case. Prior to recombination, radiation and baryonic matter interacted freely though dark matter was not involved. In addition, this will not be true in models where dark energy interacts with matter. If these interactions are weak, this approximation will hold roughly true. Then, the evolution of each component, after using the equation of state (1.6), follows

$$\rho = \rho_0 \, a^{-3\,(1\,+\,w)} \ , \tag{1.21}$$

where ρ_0 is the individual energy density component at the current time when a is defined to equal unity. The evolution of the given components follows

$$\rho_M = \rho_{M,0} \, a^{-3} \ , \tag{1.22}$$

$$\rho_R = \rho_{R,0} \, a^{-4} \ , \tag{1.23}$$

$$\rho_{DE} = \rho_{DE,0} \, e^{-3 \int_0^a a^{-1} \, da \, (1\,+\,w_{DE}(a))} \ , \tag{1.24}$$

where we left the evolution of the dark energy density in a generic form. For cosmological constant dark energy which the currently favored model, $w_{DE} = -1$ and $\rho_{DE} = \rho_\Lambda$. Simple models of dark energy, where w_{DE} is a constant, yield

$$\rho_{DE} = \rho_{DE,0} \, a^{-3\,(1\,+\,w_{DE})} \ . \tag{1.25}$$

To simplify the expansion history equations further, we can introduce the critical density at any point in time

$$\rho_C(t) = \frac{3\,H^2}{8\,\pi\,G} \ , \tag{1.26}$$

and specifically at the present time

$$\rho_{C,0} = \frac{3\,{H_0}^2}{8\,\pi\,G} \ . \tag{1.27}$$

The Friedman Equation 1.16 can then be re-written as

$$1 = \Omega_M + \Omega_R + \Omega_{DE} + \Omega_K \ , \tag{1.28}$$

with $\Omega_M{=}\rho_M/\rho_C$, $\Omega_R{=}\rho_R/\rho_C$, $\Omega_{DE}{=}\rho_{DE}/\rho_C$ and

$$\Omega_K = -\frac{k\,c^2}{H^2\,a^2} \ , \tag{1.29}$$

defining an *equivalent* density of space-time curvature. Multiplying both sides of Equation 1.28 by H, factoring out H_0 (the Hubble parameter at the current time also known as the Hubble constant) from the right hand side and substituting in Equations 1.22, 1.23, and 1.24 we can write it as

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = {H_0}^2 \left(\Omega_{M,0}\,a^{-3} + \Omega_{R,0}\,a^{-4} + \Omega_{DE,0}\,e^{-3\int_0^a a^{-1}\,da\,(1\,+\,w_{DE}(a))} + \Omega_{K,0}\,a^{-2}\right) \ , \tag{1.30}$$

with $\Omega_{M,0}{=}\rho_{M,0}/\rho_{C,0}$, $\Omega_{R,0}{=}\rho_{R,0}/\rho_{C,0}$, $\Omega_{DE,0}{=}\rho_{DE,0}/\rho_{C,0}$ and $\Omega_{K,0}{=}-\frac{k\,c^2}{{H_0}^2}$ defined at the current time. The fate of the universe can be determined from Equation 1.30. It is illustrative to write it as

$$\dot{a}^2 \ \propto \ \Omega_{M,0}\,a^{-1} + \Omega_{R,0}\,a^{-2} + \Omega_{DE,0}\,a^2\,e^{-3\int_0^a a^{-1}\,da\,(1\,+\,w_{DE}(a))} + \Omega_{K,0} \ . \tag{1.31}$$

In a world without dark energy and where radiation density can be ignored since its contribution is so small, the matter density of the universe determines its fate via

$$\dot{a}^2 \ \propto \ \Omega_{M,0}\,a^{-1} + (1 - \Omega_{M,0}) \ , \tag{1.32}$$

$$\ddot{a} \ \propto \ -\Omega_{M,0}\,a^{-2} \ . \tag{1.33}$$

The expansion of the space-time continuously decelerates but for an open universe with $\Omega_{M,0} < 1$ the expansion rate reaches a constant positive value while the deceleration tends asymptotically to 0. In this case the space-time continues to expand for ever. In a flat universe with $\Omega_{M,0} = 1$ the space-time expansion rate (and deceleration) will asymptotically approach zero but will never turn around. However, if the universe is closed with $\Omega_{M,0} > 1$ the expansion rate will reach zero in a finite amount of time and the continuous deceleration will force the space-time to start shrinking ultimately leading to a collapse. Dark energy complicates the picture. In a simple model with a constant w_{DE} (while still ignoring radiation)

$$\dot{a}^2 \propto \Omega_{M,0}\, a^{-1} + \Omega_{DE,0}\, a^{-1-3\, w_{DE}} + (1 - \Omega_{M,0} - \Omega_{DE,0}) \ , \tag{1.34}$$

$$\ddot{a} \propto -\Omega_{M,0}\, a^{-2} + (-1 - 3\, w_{DE})\, \Omega_{DE,0}\, a^{-2-3\, w_{DE}} \ . \tag{1.35}$$

If $w_{DE} < -\frac{1}{3}$ the velocity will increase and the acceleration will turn positive as the space-time expands unless the value of $\Omega_{DE,0}$ is small enough and the curvature is positive such that the universe may begin to collapse before dark energy becomes dominant. If $w_{DE} > -\frac{1}{3}$ dark energy would behave more like matter or radiation leading to a universe that may collapse depending on the curvature. In the current concordance ΛCDM model where w_{DE} = -1 the universe will keep on expanding forever at an increasing rate.

It is important to note that most sources drop the “0” symbolizing *present* time from definitions of the Ω densities. I shall follow this convention here but the reader should be aware of this and understand the context of any equation presented.

In the following section I define multiple distance measures. They require that we specify the dimensionless Hubble parameter $E(a)$

$$E(a) = \frac{H}{H_0} = \sqrt{\Omega_M\, a^{-3} + \Omega_R\, a^{-4} + \Omega_{DE}\, e^{-3\int_0^a a^{-1}\, da\, (1 + w_{DE}(a))} + \Omega_K\, a^{-2}} \ , \tag{1.36}$$

where Ω_M, Ω_R, Ω_{DE} and Ω_K are the present time values where I dropped the “0” symbol. In addition, many cosmological parameters and derived quantities scale with the Hubble constant H_0 but it is desired to use them without including its units. Hence by convention a parameter h is usually used instead and is defined as

$$h = \frac{H_0}{100\, km/s/Mpc} \ . \tag{1.37}$$

1.4.3 Cosmological Redshift and Distances

The formalism above allows for various measures of cosmological distances to be obtained, or alternatively, to use distance measurements to infer the cosmological parameters. First and foremost, we must define the cosmological redshift z. Redshift is the ratio of shift in wavelength to wavelength of a source due to its motion relative to us. For small distances in the local universe *Doppler* redshift is caused by motion of sources due to their peculiar velocities and is approximately equal to the ratio of the peculiar velocity to the speed of light. Cosmological redshift on the other hand occurs due to expansion of space itself. Light waves traveling from distant sources are "stretched" as the space-time they are embedded in expands. For the purpose of this work we will refer to cosmological redshift as simply redshift or z ignoring the Doppler effect. Redshift can be easily converted to the scale factor and vice versa

$$z = \frac{1}{a} - 1 \ . \tag{1.38}$$

We can now redefine the dimensionless Hubble parameter (Equation 1.36) in terms of the observable redshift as

$$E(z) = \sqrt{\Omega_M\,(1+z)^3 + \Omega_R\,(1+z)^4 + \Omega_{DE}\,e^{-3\int_0^z \frac{dz}{1+z}\,(1\,+\,w_{DE}(z))} + \Omega_K\,(1+z)^2} \ . \tag{1.39}$$

Coordinates (r, θ and ϕ) used in Equation 1.2 are co-moving which means that they are fixed in the local space and their relative motion is limited to the Hubble flow. The co-moving distance from Earth to a cosmological source can be computed based on its redshift. First we define the Hubble distance (though it is more of a conversion factor)

$$d_H = \frac{c}{H_0} \ . \tag{1.40}$$

The co-moving distance to a source at redshift z can be computed as

$$d_C(z) = d_H \int_0^z \frac{dz'}{E(z')} \ . \tag{1.41}$$

Of great importance are observable distances used for cosmological studies. However the co-moving distance is not sufficient to compute them if space-time curvature is different than zero since the area of the spherical shell light is projected onto is altered. For this reason we need to define the transverse co-moving distance, $d_{TC}(z)$ which accounts for this change:

$$d_{TC}(z) = \begin{cases} \frac{dh}{\sqrt{-\Omega_K}}\,\sin(d_C\,\frac{\sqrt{-\Omega_K}}{d_H}), & \text{if } \Omega_K < 0\text{, closed;} \\ d_C(z) & \text{if } \Omega_K = 0\text{, flat;} \\ \frac{dh}{\sqrt{\Omega_K}}\,\sinh(d_C\,\frac{\sqrt{\Omega_K}}{d_H}), & \text{if } \Omega_K > 0\text{, closed.} \end{cases} \tag{1.42}$$

We can now define the two observable distances used for cosmological studies. First is the luminosity distance, $d_L(z)$, derived from the naive scaling of the observed flux F relative to the intrinsic luminosity of an astronomical source L:

$$F = \frac{L}{4\,\pi\,{d_L}^2} \ . \tag{1.43}$$

The luminosity distance is a function of the transverse co-moving distance and redshift accounting for the expansion of the universe as

$$d_L(z) = (1+z)\, d_{TC} \ . \tag{1.44}$$

Another physical distance of importance is the angular diameter distance to a resolved source, $d_A(z)$ defined based on geometry as

$$\theta = \frac{s}{d_A} \ , \tag{1.45}$$

where s is the proper size of an object or structure at its location and θ is the angle it subtends. The angular diameter distance is then defined as

$$d_A(z) = \frac{d_{TC}}{1+z} \ . \tag{1.46}$$

In addition, we can define a proper distance (Weinberg 2008), the instantaneous distance between observers, using the metric as

$$d_P(r,t) = a(t) \int_0^r \frac{dr}{\sqrt{1 - k\, r^2}} \ , \tag{1.47}$$

with three solutions depending on the curvature k:

$$d_P(r,t) = \begin{cases} \sin^{-1}(r), & \text{if } k = +1; \\ r, & \text{if } k = 0; \\ \sinh^{-1}(r), & \text{if } k = -1. \end{cases} \tag{1.48}$$

1.4.4 Dark Energy

The previous sections of this work have explicitly accounted for the possibility of the accelerated expansion of the universe caused by a cosmological constant or a dynamic form of dark energy. Prior to the 1990s it was thought by most that the cosmological constant was zero and that the universe was in fact decelerating under the influence of matter. However, studies of galaxy clustering in the 1980s begun to hint at the existence of a cosmological constant. Observations favored a low Ω_M value (e.g. Maddox et al. 1990; Efstathiou et al. 1990; White et al. 1993) while inflationary theory (Guth 1981) and primary CMB tended to favor a flat universe (Smoot et al. 1992). Simulations of flat Cold Dark Matter (CDM) cosmologies also failed to reproduce reality

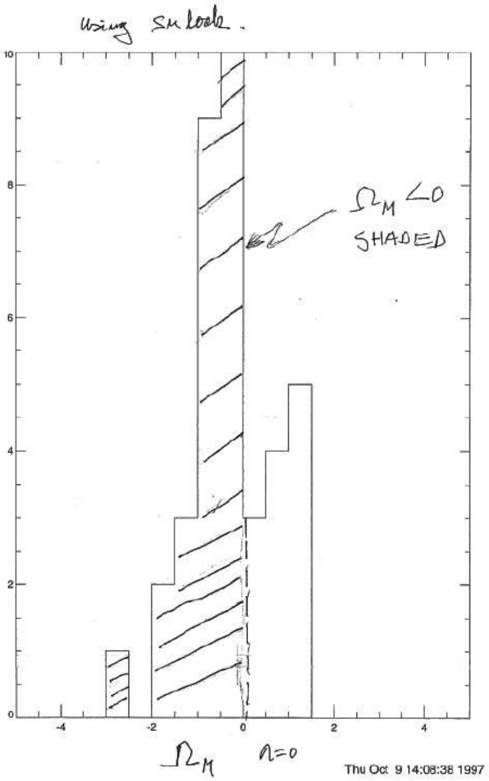


Figure 1.7: The Ω_M distribution of SNe for a universe with no dark energy. 25 out of 38 SNe observed by the Supernova Cosmology Project (Perlmutter et al. 1997, 1999) have a negative and therefore unphysical value without assuming a flat universe. Therefore, there must be another contribution to the energy density of the universe. Figure from a historical perspective discussed in (Goldhaber 2009).

(later example in Evrard et al. 2002). These arguments primed the scientific community to accept the discovery of dark energy by the *Supernova Cosmology Project* (Perlmutter et al. 1997, 1999) and *High-z Supernova Search* team (Riess et al. 1998). Figure 1.7 shows a plot of the histogram of Type Ia SNe data (from the *Supernova Cosmology Project*) vs Ω_M without assuming that the universe is flat but assuming that there is no cosmological constant. The data favor a value of $\Omega_M < 0$ which is unphysical. Allowing Ω_Λ to be larger than zero allows Ω_M to be positive. The details of the Type Ia SNe measurements will be discussed in §1.5.1.

1.4.5 Energy Conservation

A common topic of discussion and a source of confusion has emerged with the discovery of the accelerating expansion of the universe. The existence of vacuum energy seems at first glance to violate the law of conservation of energy. However that is not the case. While for a constant energy density source, the increasing volume of the universe means that more energy appears in the universe, this is balanced by the *negative* work done by the dark energy pressure. This is actually the inverse of what happens during the radiation dominated era. The radiation energy density decreases as a^4 which is faster than the volume of the universe expands, leading to an apparent loss of energy. But the loss in the total radiation energy is compensated by the *positive* work being performed by it on the universe in slowing down its expansion. Hence, if one accounts for the work being performed on space itself, energy is conserved.

1.5 Probes of Dark Energy

The discovery of dark energy ushered in a new era of cosmology. As of now, we do not understand the nature of what constitutes about 70% of the energy density of the universe. Theories as to the nature of dark energy include a cosmological constant or vacuum energy, pressure-exerting scalar fields (quintessence) or modifications to gravity (see Frieman et al. (2008) for a review). Fortunately, multiple observational techniques exist that attempt to do three things: measure more precisely the dark energy density, measure its time evolution and finally, differentiate between different theories of its source. In what follows I briefly discuss the main probes of dark energy noting that my research experience consists primarily of detector technology for use with SNe and simulations for use with galaxy cluster counts.

1.5.1 Type Ia Supernovae

Type Ia Supernovae (SNe) are the most "time-tested" probe of dark energy. They occur when a white dwarf accretes enough mass from its companion star to trigger a small nuclear reaction that leads to deflagration and possibly detonation of the entire star releasing a tremendous amount of energy. The similar lead up to and progenitor of this explosion result in a fairly consistent light output correlated with the duration of the explosion. Luminosity distance measurements based on Type Ia SNe are usually credited as the technique that led to the discovery of dark energy (Perlmutter et al. 1997, 1999; Riess et al. 1998). Low redshift SNe have previously been used successfully as part of the *distance ladder* to measure the Hubble constant H_0 (see Branch 1998) and the systematic discovery and follow up of high redshift ($z \geq 0.3$) objects allowed for the measurement of the energy density of the universe and of dark energy, then still referred to simply as the cosmological constant.

Type Ia SNe serve as *standard candles*; their intrinsic luminosity, when corrected for dust extinction, redshift, time dilation and duration (e.g. Perlmutter et al. 1997) has a dispersion of about 0.15 magnitudes which allows their luminosity distance to be calculated. Comparing the redshifts of SNe to their luminosity distances allows the expansion history of the universe to be measured. An example of various SNe

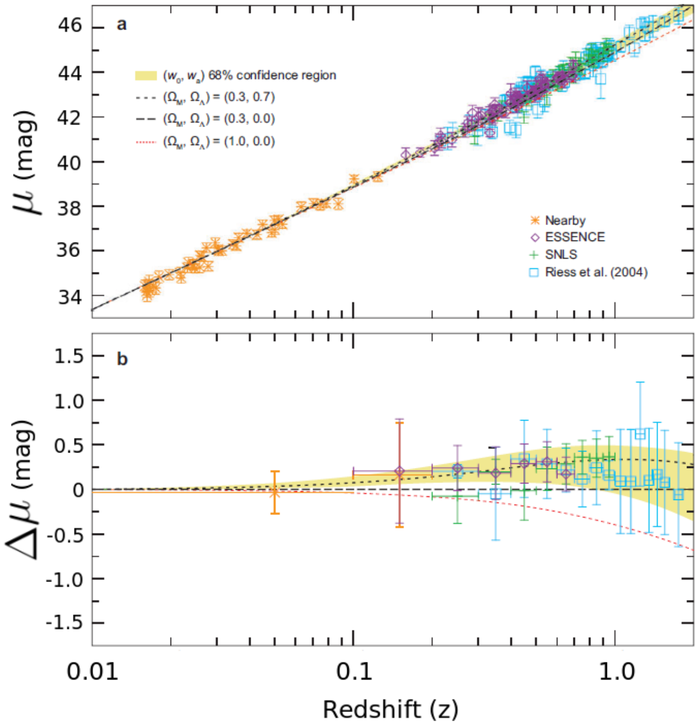


Figure 1.8: Multiple Type Ia SNe data sets. Top panel shows the distance modulus vs redshift of various data sets (symbols in the legend) along with multiple models (lines in the legend). Orange shaded region shows the $1\,\sigma$ band for dark energy model fits in a flat universe. Bottom panel shows the binned residual of the data and models from an open, matter dominated universe (long-dashed black line). The ΛCDM model (short-dashed black line) fits the data well. Figure from (Frieman et al. 2008).

data sets can be seen in Figure 1.8. It shows the distance modulus μ of SNe vs their redshift. The distance modulus is defined as as the difference between the apparent magnitude m and absolute magnitude M and is a proxy for the luminosity distance d_L normalized to 10 pc

$$\mu = m - M = 5 \times log_{10}\left(\frac{d_L}{10pc}\right) \quad . \tag{1.49}$$

Along with the data, Figure 1.8 also shows various cosmological models. The lower panel of Figure 1.8 shows the difference of the data and models for an open universe with no dark energy (long-dashed black line). The data fits the concordance ΛCDM model best (short-dashed black line).

1.5.2 Weak Lensing

Weak gravitational lensing, also known simply as weak lensing, relies on statistical measurements of the distortion of distant galaxies' shapes. Photons from these source galaxies are deflected along the way by the local curvature of space due to clumps of matter. Hence the shapes of these source galaxies are sheared in coherent ways. Observing lensed and lensing galaxies of known redshifts then allows the matter content of the universe and its evolution to be mapped. Weak lensing is a powerful probe of the nature of dark energy since the galaxy shape distortions are a function of both the expansion history and the growth history of the universe. This feature of weak lensing allows us to break the degeneracies between simple dark energy models (cosmological constant, quintessence) and models attempting to explain dark energy by modifying the theory of gravity. An example of this application (Linder 2005) is shown in Figure 1.9 where the growth history of a braneworld gravity model (Dvali et al. 2000) differentiates it from a quintessence model with an almost identical expansion history.

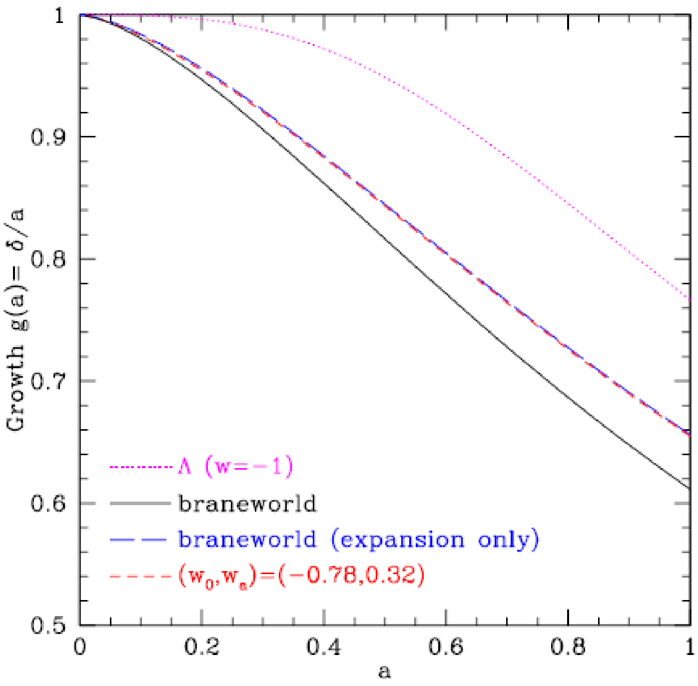


Figure 1.9: Normalized growth history of the universe vs scale factor. Growth history of the braneworld model based only on the expansion history without accounting for modified gravity (blue dashed line) is almost identical to a simple quintessence model with $w_0 = 0.78$, $w_a = 0.32$. However, the true growth history of the braneworld gravity (solid black line) as measured with weak gravitational lensing is quite different. ΛCDM cosmology is shown as a purple dotted line. Figure from (Linder 2005).

1.5.3 Baryon Acoustic Oscillation

Baryon Acoustic Oscillation (BAO) serves as an independent method of measuring the expansion history of the universe (e.g. Frieman et al. 2008; Weinberg et al. 2012) via measuring angular diameter distances (d_A) using *standard rulers* as opposed to luminosity distances using *standard candles* of SNe. Sound waves propagating through the photon-plasma medium prior to recombination imprinted a correlation in matter density on a co-moving length scale of approximately $100h^{-1}$ Mpc on the primary CMB and on the matter distribution at all epochs. This correlation can be detected in angular separations on the sky at specific redshifts as well as in line-of-sight separations using precise spectroscopic information. This allows the Hubble parameter $H(z)$ to be measured. Figure 1.10 shows a measurement of the BAO feature in SDSS data

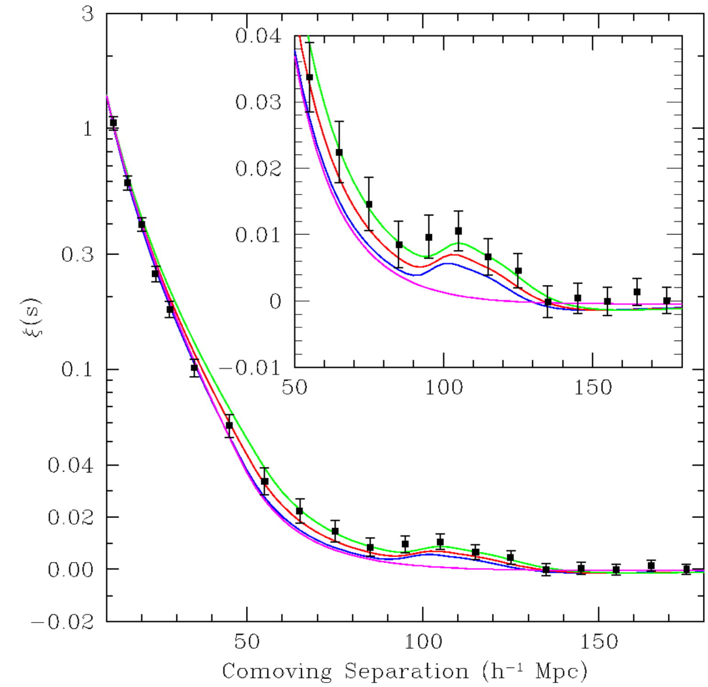


Figure 1.10: BAO peak observed in the correlation of LRGs in the SDSS data. Figure from (Eisenstein et al. 2005).

(Eisenstein et al. 2005) using luminous galaxies as traces of matter. Recently, more novel measurements have used the Lyman-α (Lyα) forest composed of absorption lines due to hydrogen gas in inter-galactic space back-illuminated by $2.1 < z < 3.5$ quasars as such a tracer. Using this technique, Busca et al. (2013) report a BAO

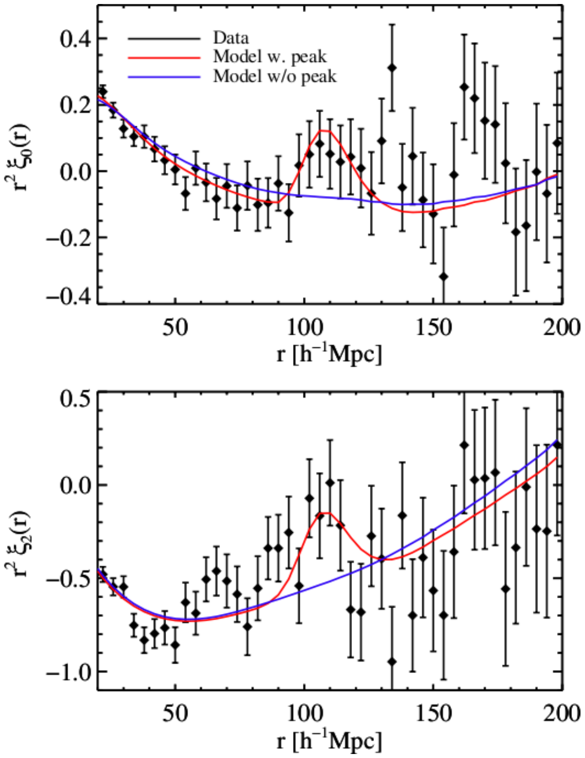


Figure 1.11: BAO peak observed in the correlation function of the Lyα forest. Top panel shows the absorption line separation^2 (angular combined with line-of-sight) × the monopole of the correlation function vs the separation. The bottom panel shows the separation^2 × the quadrupole of the correlation function. See text for the description of the monopole and quadrupole decomposition. The blue and red lines show the model fit without and with the BAO peak, respectively. Figure from Busca et al. (2013).

measurement, shown in Figure 1.11 at a median redshift of 2.3, well before dark energy became dominant. The Figure shows the monopole ξ_0 and quadrupole ξ_2 in a Legendre polynomial fit of the two point correlation function $\xi(r, \mu)$. The fit is performed in μ, the fraction of the line-of-sight separation of the total separation r at each separation

$$\xi(r, \mu) = \left[\xi_0(r) - \frac{1}{2}\,\xi_2(r)\right] + \frac{3}{2}\,\xi_2(r)\,\mu^2 \ . \tag{1.50}$$

Since μ serves as a proxy for the inverse angular separation, at each r a constant plus quadratic-in-angle model is fit to the data. Interestingly, the small secondary bump, visible in the bottom panel of Figure 1.11 at $r \approx 65\,h^{-1}Mpc$, is a strong function of μ indicating that it could be some systematic that emerges when the back-illuminating quasars appear nearly along the same line of sight. Nevertheless, this is a powerful demonstration of how BAO measurements can constrain cosmological parameters even at high redshift.

1.5.4 Galaxy Cluster Counts

Clusters of galaxies are the largest virialized (or almost virialized) structures in the universe. They trace the highest peaks in the matter density field and therefore the high end of the halo mass function (e.g. Weinberg et al. 2012). Sensitivity to cosmology comes from the geometry (co-moving volume element) and growth of structure. Figure 1.12 shows the expected number counts from a Dark Energy Survey (DES) or *South Pole Telescope* (*SPT*) like survey for different cosmological models. These calculations overestimate the cluster counts due to the high normalization (σ_8) of the power spectrum used (see §1.6.1.1 and Equation 1.63). The differences in the models are dominated by the volume element for redshifts < 0.6 and by the growth rate at higher redshifts (Frieman et al. 2008).

Because galaxy cluster abundance is exponentially dependent on mass, it is very sensitive to cosmological parameters. Its measurement as a function of mass has the potential to greatly constrain cosmological parameters (e.g. Wang & Steinhardt 1998; Haiman et al. 2001; Holder et al. 2001; Battye & Weller 2003; Molnar et al. 2004; Wang et al. 2004; Lima & Hu 2004). In addition, similar to weak lensing, galaxy cluster abundances can begin to distinguish between simple dark energy models, including the cosmological constant, and modified gravity models. Galaxy clusters can also be used as tracers of BAO and serve as matter proxies for the measurement of two point as well as higher order correlation functions and power spectra that constrain cosmology (Reid et al. 2010, for LRGs). Galaxy clusters are discussed in more detail

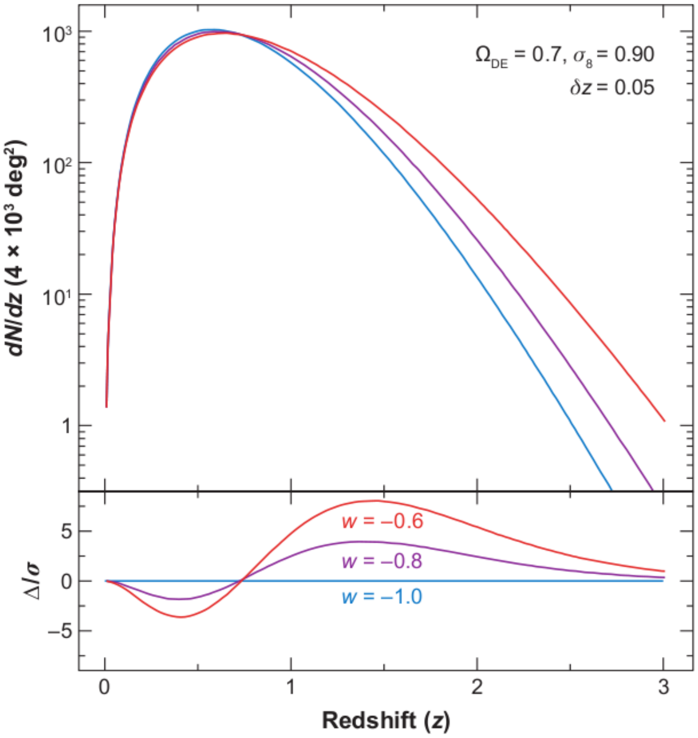


Figure 1.12: Predicted cluster counts for a DES-like or *SPT*-like survey and its sensitivity to w. Counts of clusters with mass $> 2 \times 10^{14} M_\odot$, Ω_M fixed at 0.3 and σ_8 fixed at 0.9. Bottom panel shows the difference of the models from ΛCDM . Figure from (Mohr 2005) via (Frieman et al. 2008).

in §1.6.

1.6 Galaxy Clusters

Galaxy clusters are composed of galaxies as well as hot gas that occupy the potential wells of dark matter concentrated into halos. As mentioned above, cluster abundances are used to measure cosmological parameters by comparing observed cluster densities for a given mass to halo densities predicted by N-body simulations. These simulations account for the initial conditions of the matter spectrum and its evolution subject to the force of gravity and the expansion of space-time itself. The method is limited chiefly by the proper identification of galaxy clusters, their correlation with the underlying dark matter halos and the measurement of their masses. Here, we briefly review how dark matter halos form and fill with baryons, how they are detected and "weighed" via different observables and how cosmological parameters are derived.

1.6.1 Galaxy Cluster Formation

Galaxy clusters form from regions of overdensity seeded in the primordial universe. After matter energy density becomes dominant small inhomogeneities in the distribution of pressure-less dark matter begin to collapse gravitationally (e.g. Press &

Schechter 1974; Kaiser 1991; Zentner 2007; Kravtsov & Borgani 2012) to form small halos. After recombination baryonic matter begins to fall into these halos. The growing inhomogeneities initially behave *linearly* but once they become much denser than the background they become *non-linear*. This collapse is hierarchical in nature (Press & Schechter 1974), that is, small structures form first and continue to merge as the expansion of the universe slows down. The collapse is to a large degree self-similar (Kaiser 1986) in nature as gravity does not introduce specific length scales even though baryonic physics does.

1.6.1.1 Linear Structure Growth

Dark matter halos underlying clusters of galaxies form once growth of structure becomes non-linear. First, we will describe the linear growth and the conditions under which it occurs since it will also characterize the scales at which it breaks down.

All of the structures in the current universe started out as small inhomogeneities in the matter distribution that grew linearly until their overdensity became too large. As with our treatment of Einstein's equations, we consider dark matter to be an ideal fluid that interacts only gravitationally. Initially the dark matter fluid is quite homogeneous with only small inhomogeneities, also known as the density contrast field, $\delta(\vec{x})$ defined as

$$\delta(t,\vec{r}) = \frac{\rho(t,\vec{r}) - \langle\rho\rangle}{\langle\rho\rangle} , \tag{1.51}$$

where $\rho(t,\vec{r})$ is the local dark matter density at position $\vec{r}$ and $\langle\rho\rangle$ is the mean density of the universe. t and $\vec{r}$ are the co-moving coordinates of the density contrast field. The initial perturbations are approximately Gaussian (Guth & Pi 1982; Kolb & Turner 1990) since the universe is homogeneous. As a result, they can be represented in one dimensional k-modes after a Fourier transform as

$$\delta(k) = \int d^3r\, \delta(\vec{r})\, e^{i\vec{k}\vec{r}} , \tag{1.52}$$

and they have a k-space power spectrum $P(k)$ that contains all of the statistical information available (Kravtsov & Borgani 2012) and is equivalent to the two point correlation function

$$P(k) = \langle|\delta(k)|^2\rangle . \tag{1.53}$$

Deviations from Gaussianity require higher order correlation functions to describe.

During early, linear collapse, the individual modes (or length scales proportional to $R \approx \frac{1}{k}$) of $\delta(k)$ evolve independently and at the same rate. At a later time the modes corresponding to the smallest length scales become sufficiently overdense to decouple from the linear growth and enter a period of non-linear collapse. In time, modes with larger and larger length scales begin to decouple from linear growth. These modes can be characterized (in real space) by the variance of the density contrast field smoothed with a tophat kernel $W(R)$ that defines the spherical region that is collapsing. For a region of radius R, the smoothed density field is

$$\delta_R(k) = \delta(k)\, W_R(k) \ , \tag{1.54}$$

where $W_R(k)$ is the Fourier transform of the tophat kernel $W(R)$. The variance of the smoothed density contrast field δ_R is computed from the power spectrum as

$$\sigma^2(R) = \langle \delta_R^2 \rangle = (2\,\pi)^{-3} \int P(k) |W_R(k)|^2 d^3k \ . \tag{1.55}$$

The scale R at which the variance $\sigma^2(R)$ approaches unity is the characteristic length scale of the breakdown of the linear theory. By characterizing the linear growth rate of these inhomogeneities we can learn when this collapse occurs at a given scale (when $\sigma(R)$ approaches 1) and normalize the theoretical collapse models with observable values. Hence we shall compute this growth using simplifying assumptions.

Gravitational collapse occurs when the *local* overdensity grows sufficiently. For this the use of proper coordinates τ and $\vec{x}$ instead of co-moving coordinates is warranted. We will soon switch to a co-moving frame of reference, however, to explicitly account for the expansion of the universe. Below I follow the treatment in Kolb & Turner (1990) assuming spherical dark matter collapse. This simplifies the treatment since the dark matter pressure is zero. Due to the dominance of dark matter over baryonic matter, this assumption is approximately correct. We can apply the mass continuity equation, the Euler equation of inviscid flow and the Poisson equation where the dark matter sources a gravitational potential ϕ (Equation 9.54 in Kolb & Turner (1990)) to the field in Equation 1.51. First, Equation 1.51 can be rewritten to define $\rho(\tau, \vec{x})$ as

$$\rho(\tau, \vec{x}) = \langle \rho \rangle \, (\delta(\tau, \vec{x}) \, + \, 1) \ . \tag{1.56}$$

Then, $\frac{\partial \rho}{\partial \tau} \equiv \dot{\rho} = \langle \rho \rangle \, \dot{\delta}(\tau, \vec{x})$. By dividing the continuity equation by $\langle \rho \rangle$ and since dark matter is pressure-less, Equation 9.54 in Kolb & Turner (1990) simplifies to

$$\dot{\delta} + \vec{\nabla} \cdot ([1 + \delta]\, \vec{v}) = 0 \ , \tag{1.57}$$

$$\dot{\vec{v}} + \left(\vec{v} \cdot \vec{\bigtriangledown}\right) \vec{v} + \vec{\bigtriangledown} \phi = 0 \ , \tag{1.58}$$

$$\bigtriangledown^2 \ \phi = 4 \pi \, G \, \langle \rho \rangle \, [1 + \delta] \ , \tag{1.59}$$

where $\vec{v}(\tau, \vec{x})$ is the velocity of the density contrast field $\delta(\tau, \vec{x})$ due to the Hubble flow and peculiar velocity of matter. Since the collapse is assumed to be spherical, the angular derivatives in $\bigtriangledown$ are zeros and only the radial derivative is important.

The above can be transformed to a co-moving frame $(t, \vec{r})$ and combined into a single differential equation keeping only the first order terms in the density contrast expansion. As inhomogeneities grow the approximation that $\delta < 1$ eventually breaks down. While it holds the combined differential equation is given by Equation 9.76 of Kolb & Turner (1990). This equation is reproduced below assuming pressure-less dark matter only:

$$\ddot{\delta}(t, \vec{r}) + 2H(t)\dot{\delta} - 4\pi G \langle \rho \rangle \delta = 0 \ , \tag{1.60}$$

with the "dot" now representing $\frac{\partial}{\partial t}$ (co-moving derivative) instead of $\frac{\partial}{\partial \tau}$ (proper derivative). This equation can be solved by a linear combination of decaying and growing modes. Decaying modes are of no interest to us since they do not lead to structure formation but rather decay away. The growing solution can be written as a combination of a spatial and a temporal component with $\delta(t, \vec{r}) \propto D(t)$ and $D(t)$ known as the linear growth factor. In a ΛCDM cosmology, the growth factor is

$$D(a) = \frac{5 \, \Omega_M}{2} E(a) \int_0^a \frac{da'}{[a' E(a')]^3} \ , \tag{1.61}$$

where it is rescaled to be a function of the scale factor a instead of time (Kravtsov & Borgani 2012). This growth factor is the rate of growth of the variance $\sigma^2(R)$ of the density contrast field defined in Equation 1.55 and allows us to scale this variance as

$$\sigma(R, a) = \sigma(R, a = 1) D(a) / D(a = 1) \ , \tag{1.62}$$

with a=1 the present scale factor. The past exponential increase of $D(a)$ is consistent with hierarchical growth; the overdensity at a given co-moving scale grows with time. More recently, with the domination of dark energy, the growth factor is reaching a constant value where most of the matter has been enclosed in halos. We can now introduce another important cosmological parameter, σ_8 defined as

$$\sigma_8 = \sigma(R = 8 \, h^{-1} Mpc, a = 1) \ , \tag{1.63}$$

which allows us to normalize the power spectrum since it can be measured using galaxy cluster data as seen today. Its current value in the concordance cosmological model is approximately 0.8. This tells us that at the current epoch, overdensities on scales slightly smaller than $8\,\mathrm{h}^{-1}\mathrm{Mpc}$ are entering the non-linear regime. This is not a coincidence since the $8\,\mathrm{h}^{-1}\mathrm{Mpc}$ scale was chosen because past measurements indicated it was roughly equal to unity.

1.6.1.2 Non-Linear Structure Growth

Generally, once the peaks of the overdensity field leave the linear regime we must rely on sophisticated N-Body simulations. However, some simplistic models can be constructed that provide us with intuition about the non-linear growth. For instance, a uniform overdense sphere collapsing in an Einstein-de Sitter universe ($\Omega_M = 1,\ \Omega_\Lambda = 0$) will form a halo with final virialized density contrast relative to the matter density ρ_M at the time of collapse (Kravtsov & Borgani 2012)

$$\Delta_{vir,\,m} = \langle \rho_{Halo} \rangle_{vir} / \rho_M \approx 178 \ , \tag{1.64}$$

where $\langle \rho_{Halo} \rangle_{vir}$ is the halo density average over the virial radius. This remains approximately true in a ΛCDM universe since much of halo formation occurred while the universe was matter dominated. However, as the universe expanded, the value of $\Delta_{vir,\,m}$ increased as the matter density continued to decrease.

Halo masses and radii are usually defined using a set density contrast as in Equation 1.65 relative to matter density

$$\Delta_m(z) = \langle \rho_{Halo} \rangle_R / \rho_M(z) \ , \tag{1.65}$$

or, more commonly, the density contrast relative to the critical density

$$\Delta_c(z) = \langle \rho_{Halo} \rangle_R / \rho_C(z) \ . \tag{1.66}$$

Both use the halo density average over some radius; in fact, this radius is usually chosen to make Δ a particular value. For example, a halo at a redshift z has a mass $M_{\Delta,\,c}$ and a radius $R_{\Delta,\,c}$ such that

$$\frac{M_{\Delta,\,c}}{4/3\,\pi\,R_{\Delta,\,c}^3} = \Delta \times \rho_C(z) \ , \tag{1.67}$$

and a mass $M_{\Delta,\,m}$ and a radius $R_{\Delta,\,m}$ such that

$$\frac{M_{\Delta,\,m}}{4/3\,\pi\,R^3_{\Delta,\,m}} = \Delta\ \times\ \rho_M(z)\ . \tag{1.68}$$

Note that often just M_Δ and R_Δ are used. In those cases, the text should clarify whether the density contrast Δ is relative to the mean matter or critical densities of the universe. Some common density contrasts are $\Delta_c = 500$, often used in X-ray cluster studies, $\Delta_c = 200$ common in numerical simulation descriptions and $\Delta_m = 200$ used for optical clusters though less common now. In this work Δ shall refer to the density contrast relative to the critical density $\rho_C(z)$ unless explicitly stated otherwise.

1.6.1.3 Baryonic Matter Collapse

The dark matter halos serve as sources of gravitational potential which attracts the diffuse gas from the initial, nearly homogeneous distribution. Unlike dark matter, the infalling gas is heated via compression and shocks (e.g. Kravtsov & Borgani 2012) allowing it to radiate energy and collapse further at the locations of local overdensities and eventually forming galaxies. Unfortunately for observers, the picture gets far more complicated once that happens. The formation of stars serves as a sink for cooler and denser gas changing its thermodynamic properties though most of the baryons remain in the intracluster medium (ICM). At the same time SNe explosions can now serve to heat the surrounding gas and drive winds throughout the ICM. In addition, gas accretes onto super massive black holes in centers of clusters to form Active Galactic Nuclei (AGN) (e.g. Borgani & Kravtsov 2009) which can further inject heat into the ICM (see review in McNamara & Nulsen 2007). Overall, describing the observable environment of galaxy clusters is quite complicated and requires careful numerical studies as well as the use of less sensitive proxies.

1.6.1.4 Numerical Simulations

For a full treatment of linear and non-linear structure formation we must rely on sophisticated N-Body simulations that evolve initial matter spectra into particle halos that can be compared to the halos that host galaxy clusters in the real universe. These simulations became useful once computing power became sufficient to allow for sufficiently complex structure formation codes. Figure 1.13 shows snapshots of the the dark matter distribution in the Millennium simulation (Springel & et al. 2005) on the left hand side. Semi-analytic tools can be used to populate the dark matter only simulation with galaxies as on the right hand side of that figure.

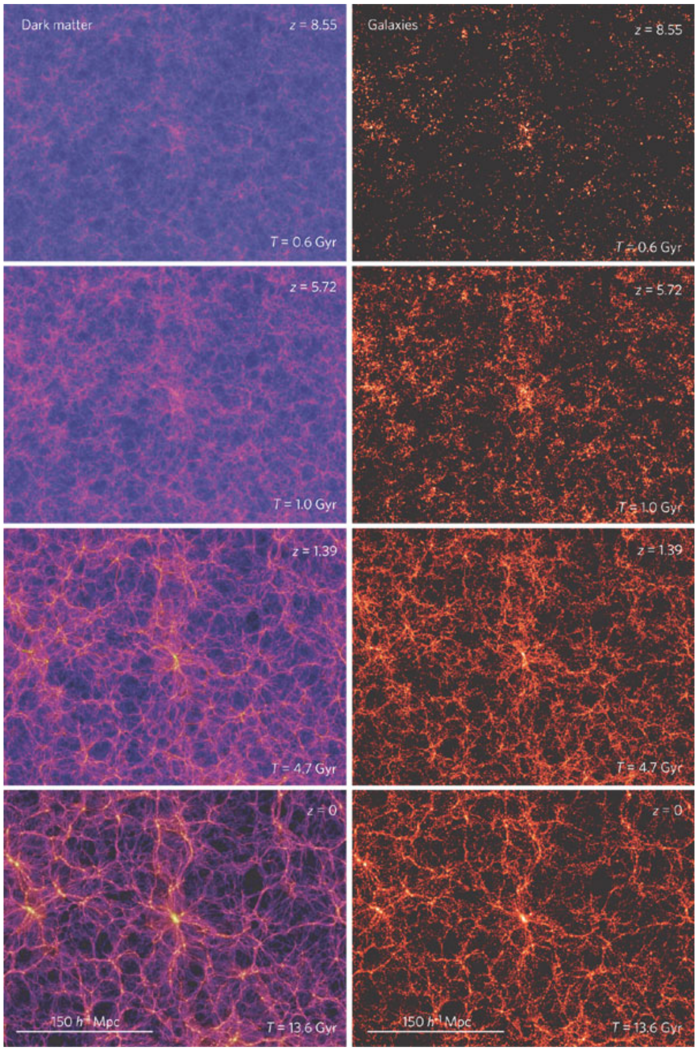


Figure 1.13: Simulated evolution of dark matter and galaxies in the Millennium simulation (Springel & et al. 2005). Left panels are the snapshots of the dark matter distribution at redshifts of 8.55, 5.72, 1.39 and 0. The color ranging from blue to purple indicates increasing velocity dispersion of simulated particles and the brightness indicates the increasing log(density) of particles. Right hand side panels show galaxy populations simulated with semi-analytic models on the underlying dark matter in the left panels. The color scale corresponds to log(stellar mass). Figure from Springel et al. (2006).

One such set of simulations is produced blindly, the Blind Cosmology Challenge (BCC) (Busha & et al. 2013; Wechsler & et al. 2013) as part of the calibration of the DES in order to prepare for full scale cosmological analysis. These products are composed of large, dark matter only simulations with galaxies and their observable quantities added using the Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS) algorithm. In addition to these optical catalogs, I have developed a framework to prepare SZ simulations using the same underlying dark matter halos for joint optical – SZ studies. This framework is discussed in Chapter 2. One of the basic outputs of the BCC dark matter simulations are the halo mass functions that are very sensitive to cosmological parameters. An example of the *Aardvark 1.0* simulated halo density is shown in Figure 1.14. It also reveals some limitations of numerical simulations. There are three obvious distinct redshift ranges in the figure

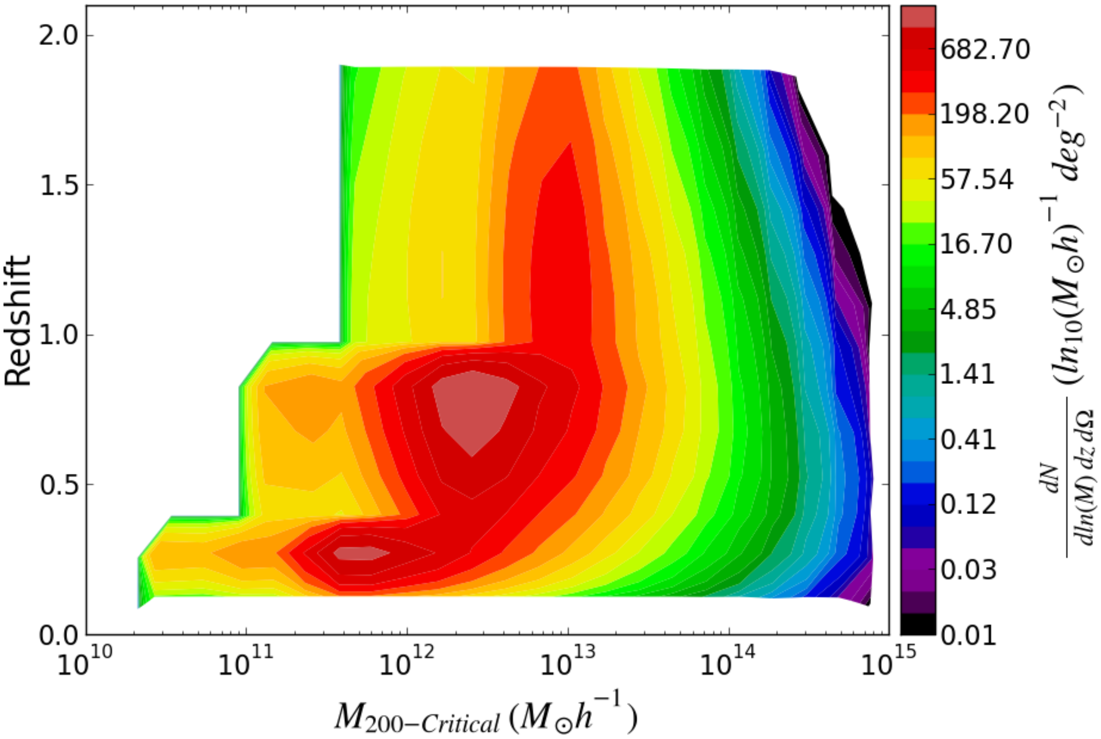


Figure 1.14: Mass function of halos in the Aardvark 1.0 simulation for DES.

caused by the fact that the simulated light cone is constructed from three N-Body runs with different resolutions. This means that the low mass limit of the mass function is different in these ranges and thus great care must be taken when studying low mass halos ($M_{200} < 10^{13} h^{-1}\, M_\odot$). The goal of galaxy cluster count studies is to essentially fit this mass function while accounting for observable effects.

Unfortunately, even though the evolution of dark matter is well understood and computationally modeled, it cannot be observed directly. Hence gas has also become an important part of N-body simulations (e.g. Voit 2005; Borgani & Kravtsov 2009). Various teams (e.g. Nagai 2006; Shaw et al. 2008; Sehgal et al. 2010; Stanek et al. 2010; Battaglia et al. 2010; Kay et al. 2012) run simulations where a fraction of the simulated particles are treated as interacting (and therefore able to be heated) and implement semi-analytic models aimed at addressing some or all of the effects discussed in §1.6.1.3. The ad-hod modeling of gas physics combined with the great variety of results (see Figure 1.15 for an example) means that these simulations are best used for setting limits and exploring the possible effects of baryon interactions than for actually making predictions. At the current stage, the goal is instead to learn more about the physics of gas interaction in clusters by matching the various simulation predictions to observations.

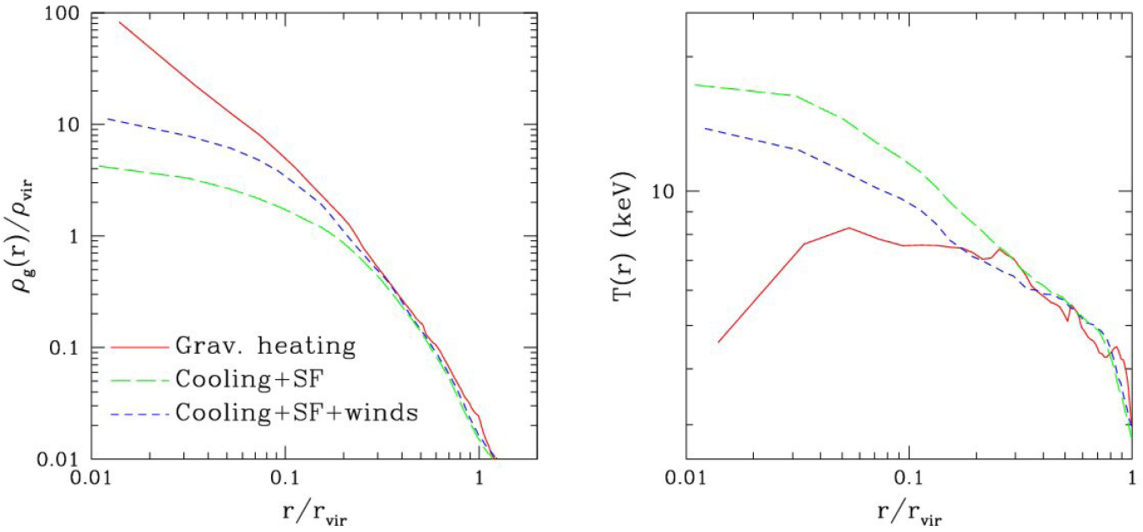


Figure 1.15: Impact of gas physics on cluster simulation. The left panels shows the gas density–virial density ratio vs radius while the right panel shows the temperature vs radius for simulations of a single cluster using three different gas physics implementations. The solid red line represents gravitational heating-only model, the long-dashed green line represents a model that also includes cooling and star formation and finally, the short-dashed blue line adds in SNe driven winds. Figure from Borgani & Kravtsov (2009).

1.6.2 Cluster Observables

Clusters of galaxies can be detected via locating galaxy count overdensities in optical or infrared (IR) observations, bright emission peaks in X-ray observations and the SZ signature in millimeter wave observations. All of these techniques allow multiple avenues for studying clusters. One of the most important features of cluster catalogs for cosmological studies is the calibration of a mass – observable relation. Different techniques can have multiple observables that vary in their uncertainty and scatter at fixed mass.

The earliest observed signatures of galaxy clusters, and their namesake, are the large number of galaxies occupying a relatively small location on the sky and in redshift space. Identifying cluster galaxies from among background galaxies is a difficult process (e.g. Gal 2006) compounded by the fact that dark matter halos themselves can be complicated structures with no clear boundaries (e.g. Behroozi et al. 2013). Various techniques will look at space densities of galaxies, their luminosities, photometric or spectroscopic redshifts and colors. Optically selected galaxy clusters can be used to correlate proxies such as aperture luminosities or optical richness (number of a type of a galaxy in a given aperture) (e.g. Koester et al. 2007a; Rozo et al. 2009b)

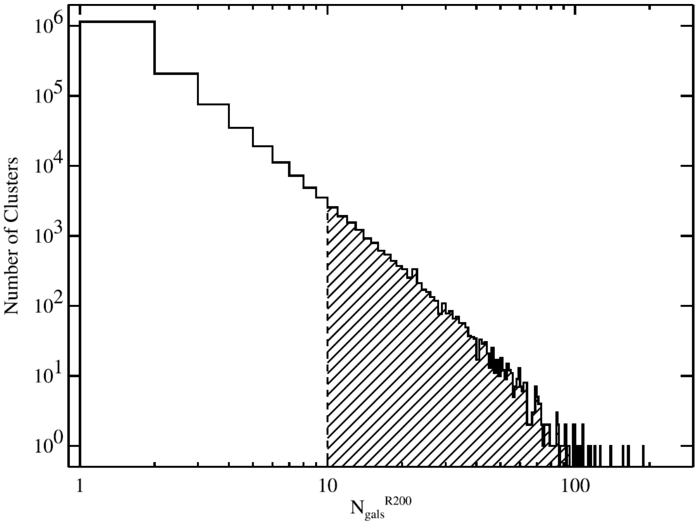


Figure 1.16: The number of objects in the maxBCG cluster catalog vs their optical richness. Line histogram includes all found objects while the shaded histogram includes objects comprising the public maxBCG cluster catalog. Note that the richness is defined as the number of bright red galaxies inside $R_{200,m}$ (see Equation 1.65 for this radius definition). Figure from Koester et al. (2007a).

to masses obtained via weak lensing (e.g. Sheldon et al. 2009; Johnston et al. 2007; Rozo et al. 2009a), X-ray luminosity and velocity dispersion measurements. Figure 1.16 shows the number of objects found by the maxBCG cluster-finder (Koester et al. 2007b) in SDSS data (line histogram) along with clusters comprising the maxBCG cluster catalog (shaded region) (Koester et al. 2007a). Cluster galaxies can be observed to high redshifts though these observations require longer integration times due to the decrease in observable flux with redshift and a shift to IR detectors due to the general redshift of the light from these galaxies. Optical observations can reach clusters of fairly low masses, approaching galaxy group scales. Unfortunately, optical cluster selection function and mass scaling remain difficult to characterize and even their uncertainties are not well known.

Another crucial observable in cosmological studies of galaxy clusters is the thermal bremsstrahlung emission by charged particles in the ICM. For low mass clusters with temperature $k_B T$ below 2 keV, emission lines dominate (see Voit (2005) for a review). Cluster X-ray measurements offer a relatively clean method of detection in that they are highly concentrated and do not suffer from projections to the extent that galaxy observations do. Given sufficient X-ray flux, the spectroscopic redshift of X-ray observed clusters can also be determined. The observed X-ray flux is a strong function of redshift, making it difficult to observe all but the most massive clusters above a z of 0.5.

One of the most important uses of X-ray clusters may be the derivation of their masses. Assuming hydrostatic equilibrium (HSE) conditions (i.e. the ICM is virialized and therefore supported in the gravitational well only by the thermal pressure

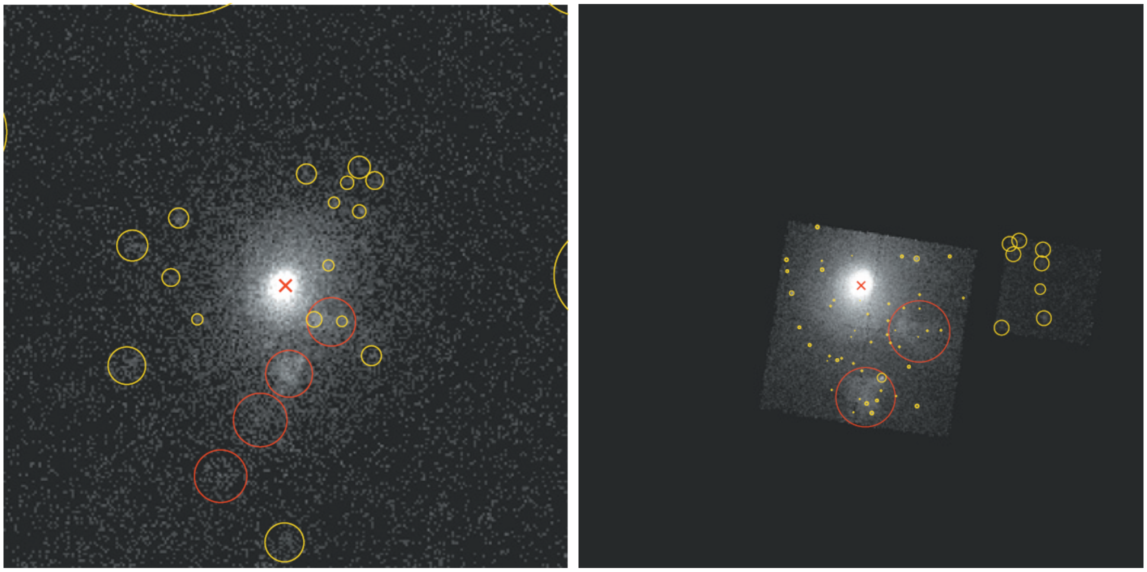


Figure 1.17: X-ray images of A85 obtained with *ROSAT* (left) and *Chandra* (right). Panels are $50' \times 50'$ in size. Yellow circles indicate additional identified sources and red circles mark substructure removed prior to image analysis. Figure from (Vikhlinin et al. 2009).

gradient) the mass M_{HSE} within a radius r is given by

$$M_{HSE}(<r) = -\frac{k_B\, T_g(r)\, r}{G\, \mu\, m_p} \frac{d}{d\, ln\, r} \left[ln\, \rho_g(r) + ln\, T_g(r) \right] \; . \tag{1.69}$$

Here k_B and G are the Boltzmann and gravitational constants, respectively, T_g and ρ_g are the three dimensional gas temperature and density profiles, m_p is the proton mass and μ is the mean molecular weight of the gas. Deep X-ray observations are needed to use Equation 1.69 since the gas density profile is obtained from well resolved X-ray surface brightness maps and the gas temperature profile is derived from the spatially resolved X-ray spectrum. Figure 1.17 shows a single X-ray cluster imaged by *ROSAT* on the left and *Chandra* on the right. *Chandra* allows for far deeper observations although its field of view is relatively small. This can make it difficult to subtract the X-ray background. Based on simulations and comparisons with weak lensing, X-ray masses of virialized objects are generally though to suffer from a negative bias of 10% to 20% (e.g. Piffaretti & Valdarnini 2008; Rasia et al. 2012). However the exact value of this bias is not yet settled on (e.g. Mahdavi et al. 2013; Planck Collaboration et al. 2013d). This bias may result from non-thermal support like bulk motion of gas and shocks, from cold gas clumps and other sub-structure or simply from the fact that the

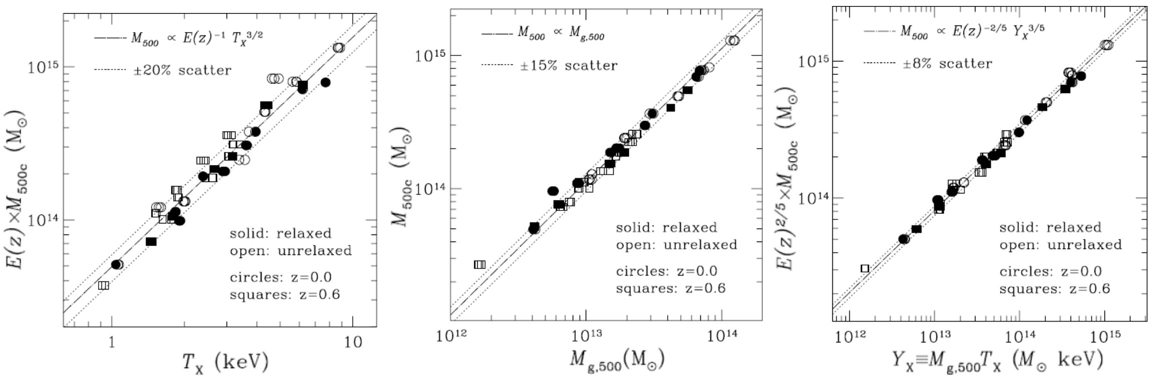


Figure 1.18: Scatters of various X-Ray mass proxies in simulations. The redshift-scaled halo mass M_{500} is shown vs aperture temperature T_X (left panel), gas mass M_g (center panel) and Y_X, a combination of T_X and M_g. Solid circles show relaxed halos and open ones show unrelaxed halos. Y_X is not sensitive to the morphological state and has the smallest scatter. Figure from Kravtsov et al. (2006).

cluster has not yet virialized and there is ongoing infall of gas. This method of direct mass computation should not be used for unrelaxed clusters and cannot be used for objects with insufficient flux. Most X-ray objects have to have a mass determined via other proxies. It is easiest to observe the X-ray luminosity, L_X, integrated within an aperture but it suffers from a very large scatter. A recently developed (Kravtsov et al. 2006) proxy, Y_X, the combination of an aperture gas mass M_g and aperture temperature T_X has been shown to be insensitive to the morphological state of the clusters and to have a low scatter as seen in Figure 1.18.

In addition to X-ray observations, the cluster gas also gives rise to the SZ effect that can be used for cluster detection and characterization. The hot ICM present within clusters of galaxies contains electrons that can inverse-Compton scatter CMB photons (Sunyaev & Zeldovich 1972; Birkinshaw 1999; Carlstrom et al. 2002). The resulting distortion of the CMB depends on the motion of the electrons relative to the photons. Two classes of the distortion can be identified: the thermal Sunyaev–Zel'dovich (tSZ) effect is the distortion due to the thermal motions of the electrons and the kinetic Sunyaev–Zel'dovich (kSZ) effect is the distortion due to the bulk motion of the electrons as the cluster moves relative to the CMB with its peculiar velocity. The tSZ signal dominates the kSZ signal and is often simply referred to as SZ. The shift of the black body curve due to the tSZ effect is shown in Figure 1.19. The two effects can be distinguished by their frequency dependence; the tSZ effect causes a decrement in CMB observations below 218 GHz and an increment above while the kSZ effect has no spectral dependence. These frequency responses

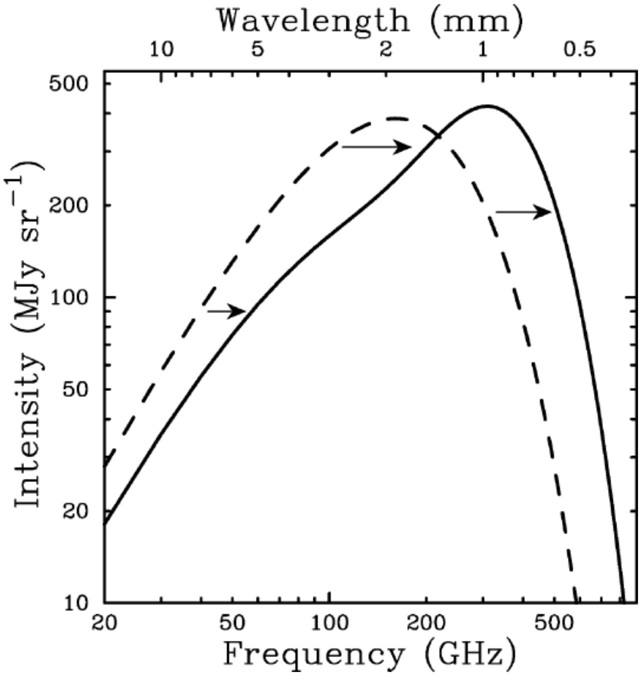


Figure 1.19: The tSZ spectral distortion effect on the primary CMB spectrum. Distorted spectrum shown as a solid line and the undistorted one shown as a dashed line. This size of the effect shown here is vastly exaggerated, the halo mass needed to achieve it is a thousand times greater than typical cluster masses. Figure from (Carlstrom et al. 2002).

are shown in Figure 1.20.

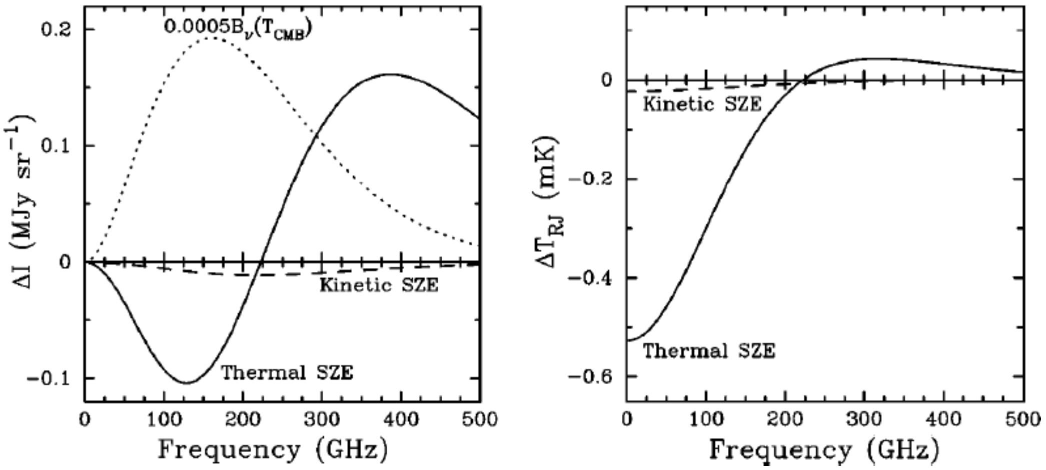


Figure 1.20: The relative effect of the tSZ and kSZ effects on the primary CMB spectrum. The plots show the difference relative to the primary CMB. The dotted line in the left panel shows a scaled primary CMB spectrum for reference. The left panel shows the source intensity and the right panel shows the thermal distortion. The solid line shows the tSZ spectrum and the dashed line shows the kSZ spectrum. This figure visualizes a massive and fast moving cluster. Figure from (Carlstrom et al. 2002).

A strong advantage of the SZ as a method for cluster detection is its insensitivity to redshift; even high redshift clusters can be detected relatively easily. However, this also leads to the problem of projection effects since high redshift background will be indistinguishable for a given cluster. In addition, current large area surveys have a fairly high mass limit of approximately $4 \times 10^{14} h^{-1}\, M_{\odot}$ due to observational noise. A simulated cluster image as it would be seen by the *SPT* is shown in Figure 1.21 at 95 GHz, 150 GHz and 220 GHz. Note that the cluster is not visible at 220 GHz where the tSZ spectral distortion is very nearly null.

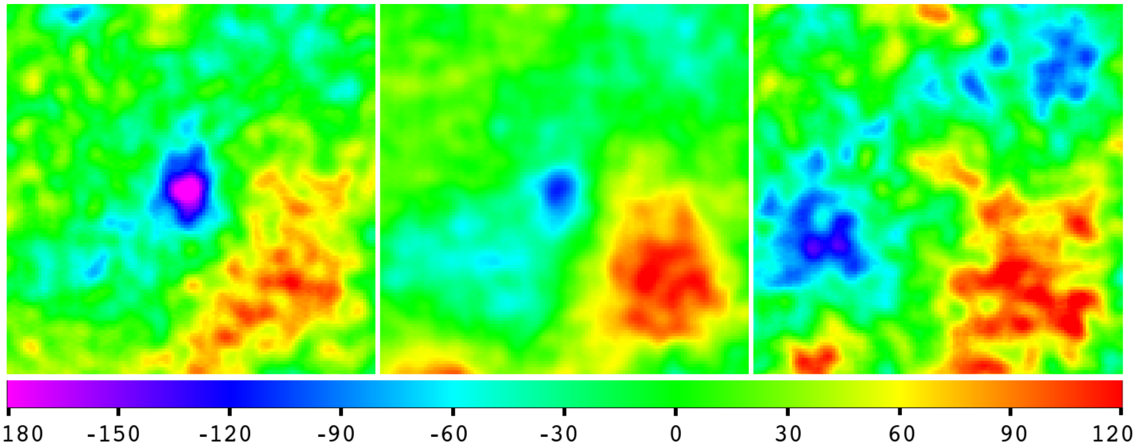


Figure 1.21: Example of a massive simulated SZ cluster as seen by the *SPT* at 95 GHz, 150 GHz and 220 GHz, left to right. The scale is in μK. The cluster has $M_{200} = 8.823 \times 10^{14} h^{-1}\, M_{\odot}$ and is located at $z = 0.163$. The images are 0.5 degree × 0.5 degree in size and are smoothed with a Gaussian kernel of $1.25'$ width. Simulated instrument noise is largest at 220 GHz and smallest at 150 GHz.

As discussed above, the three cluster detection and characterization techniques are powerful on their own, though they suffer from different systematics. By utilizing all of them in a multi-wavelength fashion it is possible to significantly improve cluster cosmology constraints (e.g. Cunha 2009; Rozo et al. 2009a; Wu et al. 2010).

1.6.3 Cosmology From Cluster Counts

All of the three cluster detection techniques have yielded cosmological parameter constraints via the method of fitting the halo mass function. This requires that various systematic effects be dealt with. These include the mass and redshift dependent cluster selection function, catalog completeness (fraction of dark matter halos identified as clusters) and purity (fraction of false detections) and most importantly, mass – observable scaling. Figure 1.22 shows the cluster counts in the maxBCG catalog with weak lensing calibrated masses (Rozo et al. 2009a). While the data (open diamonds)

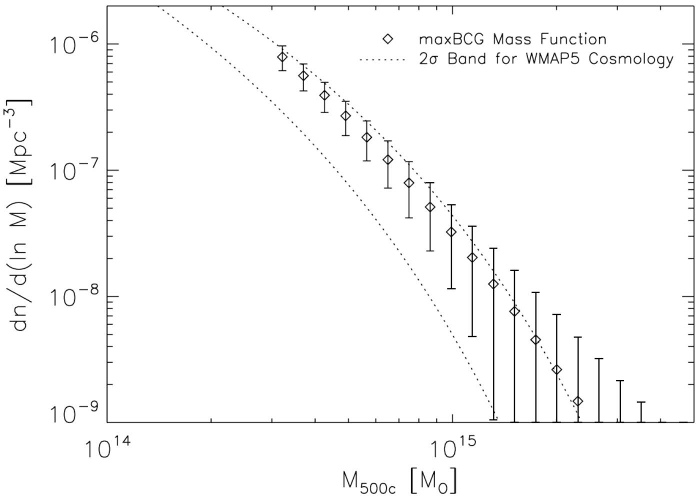


Figure 1.22: macBCG cluster counts vs mass along with ΛCDM models. The data (diamonds) are plotted along with a 95% confidence interval of concordance ΛCDM cosmology (dotted lines). Figure from Rozo et al. (2009a).

assume a ΛCDM cosmology for conversion to spatial density, they would be inconsistent with the ΛCDM mass function models (dotted lines) if the universe deviated significantly from ΛCDM .

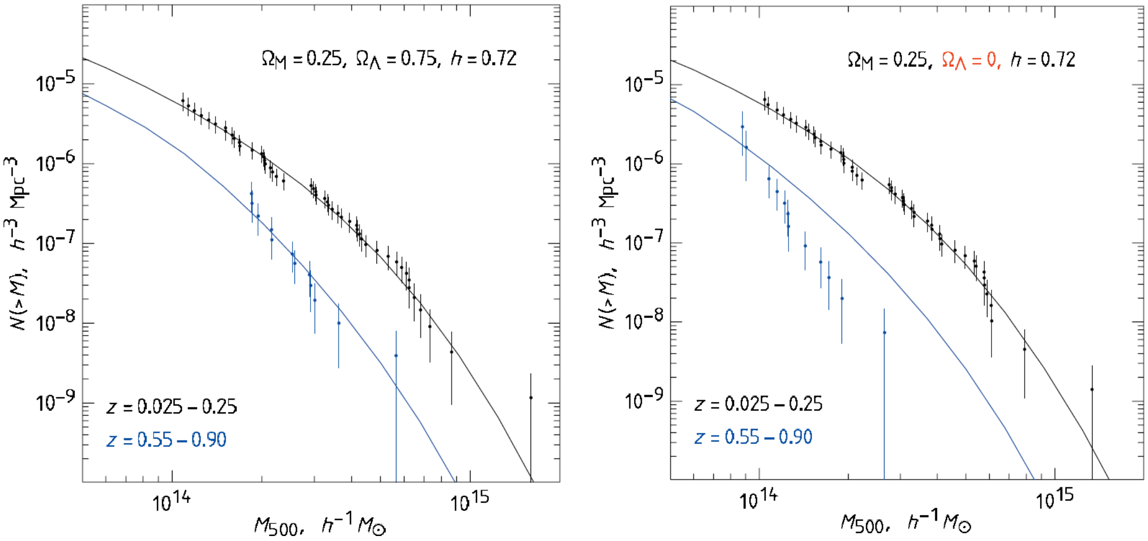


Figure 1.23: Dark energy constraints from X-ray observations. Low redshift data is shown as black dots and high redshift data as blue dots. Low redshift model mass functions are shown as black lines and high redshift models are shown as blue lines. The left panel shows data scaled assuming ΛCDM and ΛCDM mass function models. The right panel shows the data and model for an open universe with Ω_Λ=0. The mass function has been normalized at low redshift and hence the black points and black lines agree in both panels. Of note is the fact that in the left panel (ΛCDM cosmology) the blue line agrees with the blue points (high redshift data) while in the right panel (non–ΛCDM cosmology) the blue line fails to match the blue data points. Figure from Vikhlinin et al. (2009).

This is also beautifully demonstrated with X-ray data that extend towards larger redshift as seen in Figure 1.23. The low (black points) and high (blue points) redshift mass function is shown assuming a ΛCDM cosmology (left panel) and an open universe with no dark energy (right panel). The mass functions are normalized at low redshift. It is obvious that the high redshift data for the open universe disagrees with the model.

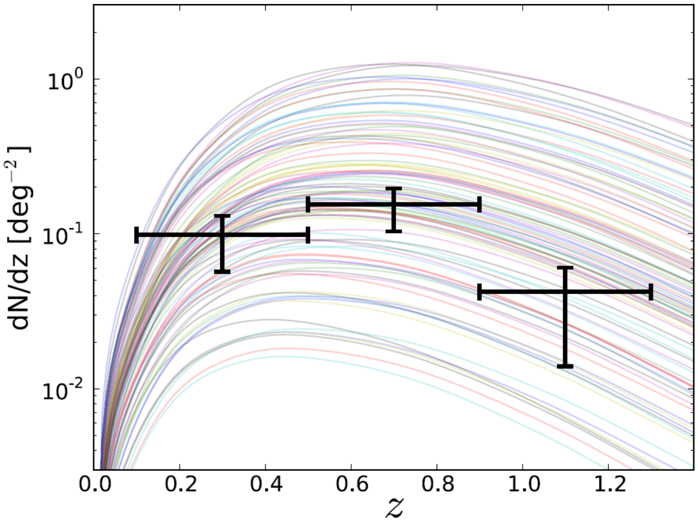


Figure 1.24: *SPT* clusters constraining the dark energy equation of state. A hundred WMAP-7 models allowing for variation of the dark energy equation of state w are shown as lines with *SPT* data shown as black error bars. Figure from Vanderlinde et al. (2010).

SZ cosmological analyses are a bit younger than X-ray and optical studies but have nevertheless be able to deliver parameter constraints. The results of one such early study (Vanderlinde et al. 2010) are shown in Figure 1.24. Here, the cluster counts derived from *SPT* observations of the SZ effect are shown as as black error bars along with a hundred realizations of the WMAP-7 cosmology allowing for the dark energy equation of state, w, being different from -1. It narrows the allowed w values by a small but certainly noticeable amount.

More recently, attempts at a joint analysis using the different galaxy cluster observational and characterization techniques have been attempted (e.g. Benson et al. 2013) though discrepancies between these methods have also been observed (Planck Collaboration et al. 2011d; Draper et al. 2012) and must be resolved (e.g. Biesiadzinski et al. 2012; Angulo et al. 2012; Rozo et al. 2012) before they can be used to their full extent.

1.7 Investigating Cosmology with Large Surveys

Large area surveys of the sky extending to high redshifts offer many possibilities for studies of the dark energy equation of state and other cosmological parameters. Such surveys will provide a large number of SNe in a variety of environments allowing for reduced statistical uncertainties and studies of systematics involved in standard candle cosmology. They will allow a better statistical measurement of BAO as well as its temporal evolution. They will also lead to improved statistics for weak lensing measurements as well as extending them to higher redshifts. Finally, a large number of galaxy clusters observed at multiple redshifts will allow us to trace the growth of structure to test for dark energy and modified gravity. In addition, it takes only a single massive, high redshift cluster to falsify ΛCDM (Mortonson et al. 2011) but finding it, if it exists, will require a large sky coverage.

1.7.1 Ground Based Optical Surveys

Optical surveys with a large coverage of the sky have already contributed greatly to our understanding of the universe, including cosmology. The SDSS alone has lead to cosmological constraints via multiple methods: clusters of galaxies (Rozo et al. 2010), BAO (Eisenstein et al. 2005; Percival et al. 2010), clustering of matter as traced by LRGs (Reid et al. 2010) and SNe (Kessler et al. 2009), just to name a few.

A new ongoing survey, the Dark Energy Survey (The Dark Energy Survey Collaboration 2005) based at the Cerro Tololo Inter-American Observatory (CTIO) in Chile, has gone through a science verification phase during the winter of 2012-2013 and will begin science operations in the fall of 2013. It promises to further push our understanding of cosmology by extending the various techniques to larger redshifts and fainter sources. Already, it is contributing to this effort with its first detection of a SNe (Abbott et al. 2012). It will also contribute a great deal to ancillary science like studies of the Milky Way, galaxy evolution and quasars. DES is planned to cover a roughly 5000 sq. degree region of the southern sky using the 4 meter Blanco telescope. The full area will be observed through g, r, i, z and Y filters ten times, twice a year, in order to reach an apparent magnitude limit of 25 in the r band. The response of the DES camera including the CCD efficiency and filter transmission is shown in Figure 1.25 for all of the filters. As part of the project, the DES collaboration constructed a new instrument called Dark Energy Camera (DECam) using $\approx$70 thick CCDs sensitive out to 1050 nm in the NIR. This sensitivity in the NIR makes it easier to peer at galaxies and other sources at larger redshifts. For example, many cluster detection and photometric redshift algorithms rely on the so called *4000 Angstrom Break* (e.g. Dressler & Shectman 1987), a spectral feature present in many cluster member galaxies. This feature moves towards redder values at higher redshifts; it occurs at 800 nm at $z = 1$. Locating clusters at that redshift therefore requires good

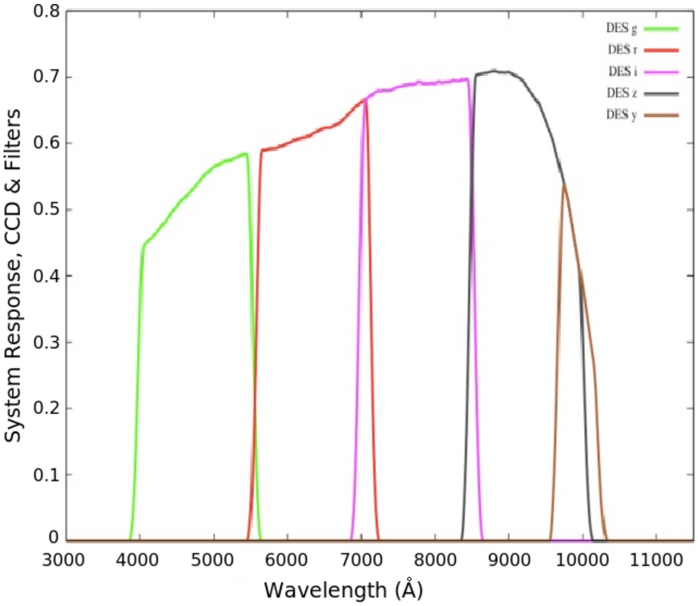


Figure 1.25: DES detector and filter response. This figure shows the response of the DES camera vs wavelength. In includes the response of the camera CCDs combined with the filter transmission curves, g in the green, r in the red, i in the purple, z in the black and Y in the brown. Atmospheric transmission is not included. Figure from (Flaugher 2011).

sensitivity in the i and z and pushing that to $z > 1.1$ necessitates the use of the Y filter. The response of the SDSS CCDs was much poorer in the z band due to the thinner detector material used. Extending the redshift coverage even further requires the use of NIR sensitive devices composed of material other than Silicon (Si). Such devices are discussed in §1.8 and their characterization constituted a large part of my graduate work as discussed in Chapters 5 and 6.

Another ongoing survey is the VISTA Hemisphere Survey (VHS) (McMahon 2010, 2012), performed in the NIR by the *Visible and Infrared Survey Telescope for Astronomy* (*VISTA*) located at Cerro Paranal, the European Southern Observatory (ESO). VHS is conducting both galactic as well as cosmological studies and overlaps with many other optical surveys, including DES. *VISTA* is a 4.2 meter telescope with a 1.5 degree field of view. Its NIR camera employs 16 HgCdTe detectors similar to those planned for *SNAP* and introduced in §1.8.

Other large, ground based surveys are planned in the near future. This includes the photometric observations performed by the *Large Synoptic Survey Telescope* (*LSST*) (Ivezic et al. 2008) located at Cerro Pachón in Chile. It is designed with an 8.4 meter mirror (with effective diameter of 6.7 meters) and a 9.6 sq. degree field of view. It will observe 30,000 sq. degrees of sky every three nights for an expected ten year operational lifetime starting around 2020. This large area coverage and cadence will enable very precise cosmological studies.

In addition to the photometric surveys mentioned above, a great deal of effort has been put into large scale spectroscopic surveys by the community. The original SDSS itself had integrated spectroscopic follow up of many of its sources. The currently ongoing SDSS-III includes the Baryon Oscillation Spectroscopic Survey (BOSS) which measures angular diameter distances and the expansion history of the universe via BAO using galaxies as well as the Lyα forest as tracers of matter at high redshifts (Busca et al. 2013). After BOSS completes its survey its instrumentation will be reused for the Extended Baryon Oscillation Spectroscopic Survey (eBOSS) as part of SDSS-IV. eBOSS will target LRGs, Emission Line Galaxies (ELGs) and quasars in between the BOSS redshift ranges and will continue studying Lyα forests. Both BOSS and eBOSS will also study redshift space distortion (RSD), small signatures of peculiar velocities in spectra of galaxies that trace the matter content and therefore, the growth of structure in the universe (e.g. Weinberg et al. 2012). A little further in the future, the Mid-Scale Dark Energy Spectroscopic Instrument (DESI) based at the 4 meter Mayall telescope at Kitt Peak in Arizona will survey a 14,000 to 18,000 sq. degree area of the sky measuring BAO and RSD to unprecedented precision as well as helping to constrain the sum of neutrino masses (Weinberg et al. 2013).

1.7.2 Space Based Optical Surveys

Optical and NIR surveys in space are in many ways similar to the ground based surveys discussed above. However they do have some unique advantages as well as problems that must be addressed.

Experiments in space have the advantage of avoiding Earth's atmosphere. This aids in two ways; lower photon background, especially for NIR observations, and a stable point spread function (PSF). NIR observations are crucial for studying the possible evolution of dark energy since light from galaxies and SNe, for example, is shifted towards that part of the spectrum if the sources are at higher redshift. Earth's atmosphere blocks portions of the NIR due to water absorption, decreasing the available flux. More importantly, atmospheric OH emission lines add to the overall background via a sky glow. This background leads to a reduction in the signal to noise ratio of astronomical sources and makes it very difficult to perform NIR spectroscopy. While narrow blocking filters are under development to reduce the sky glow (e.g. Günster et al. 2011), going to space completely removes it along with opening previously blocked regions of the electromagnetic spectrum. In such a case the limiting backgrounds comes from zodiacal light, sunlight reflected from dust grains within the solar system. The zodiacal light is up to 500 times dimmer than the NIR sky background (Brown 2007). In addition, the atmosphere tends to blur light leading to an increase in the apparent PSF of instruments. The size and stability of the PSF is of utmost importance to studies of weak lensing since it can drastically alter the shapes of source galaxies. Ground based observatories are limited to an angular resolution no better than $\approx 0.5''$ without the use of adaptive optics. Space based observatories can achieve PSFs smaller than $0.1''$ and reach their diffraction limit in the NIR. Varying mechanical stresses impacting large telescopes on the ground can also degrade the PSF but are of no concern in microgravity. Finally, space based observatories are not subject to the diurnal cycle and weather and can operate continuously for extended periods of time. All of these reasons make space based observatories ideal for surveys of the cosmos. There are also disadvantages involving space missions. The chief one is the cost of developing such missions due to the required instrument reliability and redundancy as well as launch costs. This often necessitates compromises in satellite design. For instance, it is cost-prohibitive to launch a very large focal plane into space which means that either the field of view will be limited (not practical for a survey telescope) or that the plate scale will be large leading to undersampled observations. This issue can be addressed with extensive instrument characterization and calibration but it does require quite a bit of advance planning.

There are several space survey missions planned for the near future (Weinberg et al. 2013). One scheduled for launch in 2020 is the ESA *Euclid* dark energy satellite designed to study growth of structure via weak lensing and RSD and the expansion history of the universe via BAO. It will also help lower the limit of the sum of neutrino masses. It is a 1.2 meter telescope with a $0.1''$/pixel plate scale, equipped with a visible and a NIR imager and a NIR spectrograph. It will conduct a 15,000 sq. degree, single band visible and 3 band NIR photometric survey for weak lensing measurements (visible observations for the shear and NIR observations for photometric redshifts) and a 15,000 sq. degree spectroscopic survey for BAO and RSD measurements. Scheduled to launch three years after *Euclid* is NASA's *Wide Field Infrared Survey Telescope* (*WFIRST*). While the mission specifics are still under development, it is likely to be a 2.4 meter telescope with a plate scale of $0.11''$/pixel and a 0.28 sq. degree field of view. Similarly to *Euclid*, *WFIRST* will conduct measurements of weak lensing, BAO and RSD though it aims for a smaller area but deeper survey. The biggest difference is that *WFIRST* will also conduct a deep SNe survey extending to redshifts of 1.7. While SNe are not currently in vogue as far as probes of dark energy are concerned, it remains one of the most tested and reliable techniques. Note that both *Euclid* and *WFIRST* require extensive calibration due to their use of Mercury-

Cadmium-Telluride (HgCdTe) detector technology which makes them susceptible to effects like reciprocity failure. As part of my graduate career I studied these types of detectors extensively and describe their reciprocity failure as well as their complicated quantum efficiency in Chapters 5 and 6 (also see Appendix B). In addition, *WFIRST* is an undersampled telescope, its full width at half maximum (FWHM) at 1.2μm wavelength is 0.1$''$, the same size as its plate scale. Meanwhile *Euclid* could be called "marginally" undersampled since its FWHM at 1.2μm wavelength is 0.21$''$, twice its plate scale. This makes both telescopes susceptible to intra-pixel variations (Barron et al. 2007) and also, see Appendix C.

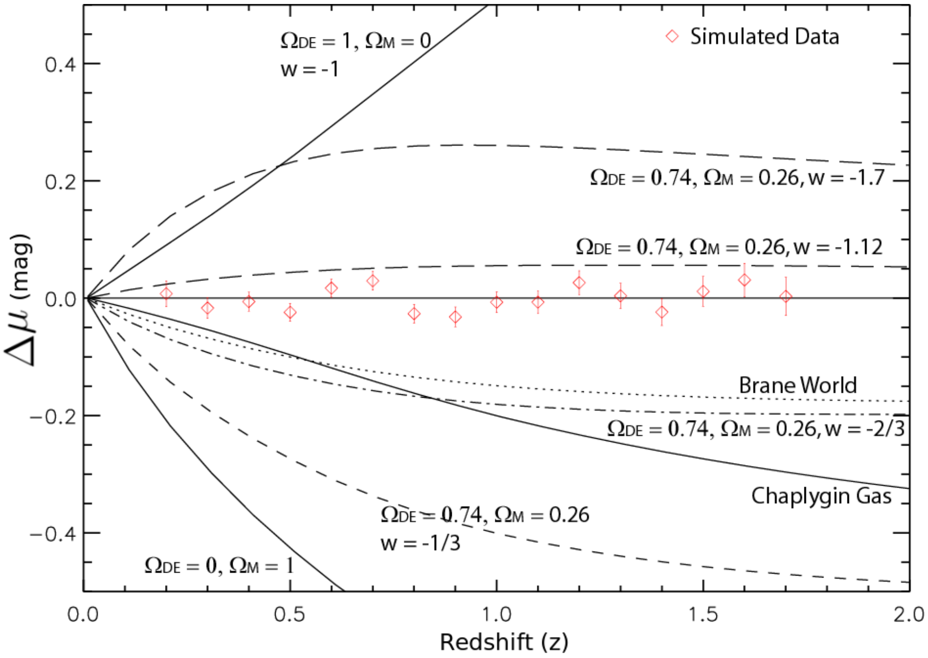


Figure 1.26: Expected discrimination of different cosmological models using SNe data from *SNAP*. Comparing to current data sets in Figure 1.8 (where the bottom panel shows magnitude difference ranging from -1.75 to 1.75) one can see the incredible improvement in the ability to distinguish between various cosmological expansion histories of the universe.

WFIRST evolved from the previously defined *Joint Dark Energy Mission* (*JDEM*) which itself was based on the *Supernovae Acceleration Probe* (*SNAP*) mission (Aldering et al. 2002). My studies of HgCdTe devices were conducted as part of the *SNAP* development. This mission was designed to take advantage of SNe, BAO and weak lensing (SNAP Collaboration 2005) measurements in visible as well as NIR wavelengths. Figure 1.26 shows the incredible measurement precision of *SNAP* and its ability do discriminate between different cosmological models with SNe data alone.

Though delayed, the combined power of *Euclid* and *WFIRST* should make such precise measurements a reality.

1.7.3 SZ Surveys

In addition to the optical surveys discussed above, the SZ effect has become an important part of cosmological studies. Clusters found in large SZ surveys have several advantages over those detected in the optical. The SZ flux is generally thought to correlate tightly with mass (e.g. Shaw et al. 2008) which improves mass estimates. It also means that SZ cluster-finding is not as sensitive to the Malmquist bias that emerges by making selection cuts on a mass proxy instead of mass itself. In addition, the background of SZ surveys is quite simple, nearly Gaussian for current generation instruments, though it becomes more complicated as the instrument noise decreases. Figure 1.27 compares the probability density of pixels being clusters for *SPT*-like simulations and *SPT-POL*-like simulations. *SPT-POL*, an *SPT* upgrade that targets CMB polarization, will have a noise level about three times lower than *SPT* while maintaining a similar beam profile. For *SPT*, a simple cut at a signal to noise (S/N) = 5 provides a very clean cluster sample and a cut of 7 would theoretically eliminate false detections altogether. *SPT-POL* would require a S/N=15 to achieve a similar false detection rate. Of course *SPT-POL* will detect more clusters and hence a given false detection rate does not degrade *SPT-POL* purity as much as *SPT*. Nevertheless, this issue must be considered as experiments improve. Another great advantage of SZ surveys is their relative insensitivity to cluster redshifts. CMB distortions do not get dimmer with increasing redshift as is the case for optical and X-ray measurements. This, combined with the simple backgrounds, makes the SZ cluster selection function very “clean.” The low mass limit of cluster-finding as well as its completeness is a relatively constant function of redshift and simple to estimate after certain assumptions about the mass–SZ scaling relation are made. Unfortunately, the lower mass limit of detectable clusters is significantly higher than for optical surveys, though it is improving. In addition, survey instruments may not be able to resolve SZ substructure (Komatsu et al. 2001) that may be of importance.

Since the first SZ clusters were blindly identified at the end of the last decade (e.g. Staniszewski et al. 2009) multiple surveys have been undertaken that significantly increased the number of known SZ clusters. These include the *Atacama Cosmology Telescope* (*ACT*) located at Cerro Toco in the Atacama Desert (Swetz et al. 2011). It is a 6 meter telescope equipped with a transition edge sensor (TES) camera to

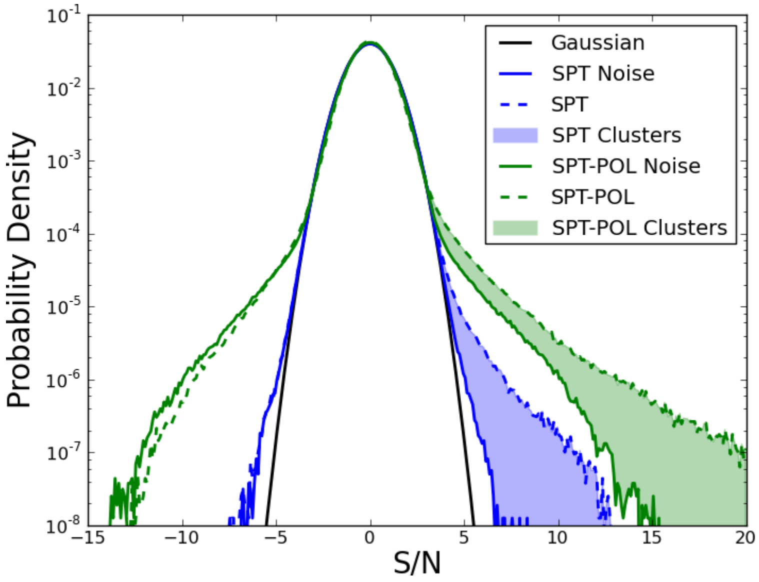


Figure 1.27: Probability density of pixels being SZ clusters plotted vs the pixel S/N ratio. Random Gaussian field example is shown as a black line. The *SPT* noise-only probability density (solid blue line) deviates slightly from the Gaussian example. *SPT* pixels (dashed blue line) have a significant tail due to presence of clusters (blue shaded region). The *SPT-POL* noise-only probability density (solid green line) deviates significantly from the Gaussian example. *SPT-POL* pixels (dashed green line) also have a significant tail due to the presence of clusters (green shaded region) though a S/N cut would not be as clean there.

observe the sky at 148 GHz, 218 GHz and 277 GHz. It has successfully found many clusters (Marriage et al. 2011, e.g.) at a resolution of roughly 1.4′ and published cosmological results (Sehgal et al. 2011). It will be upgraded in the near future to study the polarization of the CMB.

In addition to *ACT*, the *Planck* satellite has carried out a cluster survey (e.g. Planck Collaboration et al. 2011b) in the SZ as part of its much larger mission. Its resolution is considerably poorer, between 4.5′ to 10′ depending on the observational band, which limits it to lower redshift clusters. However it does have the unique advantage of a full sky coverage.

The SZ survey most applicable to my work is that carried out by the *SPT* collaboration. *SPT* is a dedicated telescope (shown in Figure 1.28) situated at the South Pole constructed to study the primary CMB, clusters of galaxies, point source radio emissions and other phenomena. It does this by studying arcminute – sized CMB anisotropies (e.g. Carlstrom et al. 2011; Staniszewski et al. 2009) over a 2500 sq. degree patch of sky visible from the South Pole that overlaps with the DES optical survey. It is a 10 m off-axis telescope and its focal plane is populated by TES bolome-

Figure 1.28: The *SPT* under construction. Image credit: Steve Padin

ters that can observe the sky at approximately 95 GHz, 150 GHz and 220 GHz bands. *SPT* has completed a survey of 2500 sq. degrees in 2011 and the collaboration is busy producing science results. It has been able to explore a variety of topics in modern cosmology including clusters of galaxies (e.g. Vanderlinde et al. 2010; Williamson et al. 2011; Reichardt et al. 2013), primordial element abundances and cosmological parameters (e.g. Keisler et al. 2011) and CMB lensing (Keisler et al. 2011; Bleem et al. 2012). The telescope itself has been reconfigured as a polarization experiment *SPT-POL*. The overlap in observed area between the *SPT* and DES makes it possible to conduct joint cosmological analyses that may result in significantly improvements over their independent results.

1.8 NIR Detectors

The development of visible-light digital sensors changed the nature of astronomy permanently by making it possible to observe larger areas of the sky and analyze the gathered data faster. Developments of IR focal plane arrays (FPAs) have enabled astronomers to pear through dust and observe the far reaches of the universe. As discussed in §1.7.2 the *SNAP/JDEM* mission planned on utilizing NIR detectors in order to observe galaxies and Type Ia SNe at high redshift and use them as tracers of cosmic expansion and growth of structure. The *Euclid* telescope will also use such devices for its BAO survey and the VHS is currently using them from the ground. While CCDs are usually constructed from Si, NIR detectors come in a multitude of compositions. *SNAP* investigated HgCdTe and indium gallium arsenide (InGaAs) based FPAs for its use. At the time, InGaAs technology was limited by its temperature dependent bandgap and a general lack of development and hence the more mature HgCdTe composition was selected (Brown 2007).

The detector characterization laboratory at the University of Michigan was charged with testing the performance of engineering devices and providing feedback to the manufacturers, Teledyne Imaging Sensors (TIS). As part of this work, the lab along with our collaborators has studied the "traditional" effects such as dark current, read noise and quantum efficiency (QE) as well as less known effects like pixel self heating, capacitive coupling, persistence and intrapixel variation (e.g. Brown 2007; Schubnell et al. 2006, 2008; Brown et al. 2006; Smith et al. 2008; Barron et al. 2007). As part of my graduate work I helped to perform additional novel detector studies in the characterization lab including studies of reciprocity failure (see Lorenzon et al. 2008) discussed in Chapter 5, multi-parameter QE characterization discussed in Chapter 6, large scale detectors sensitivity variations (Appendix B) and detector wide studies of intrapixel structure (Appendix C) including pixel size variation (e.g. Smith & Rahmer 2008; Lorenzon et al. 2008). Below I shall discuss the use of HgCdTe FPAs in greater detail.

1.8.1 HgCdTe Detectors

Mercury-Cadmium-Telluride allows for tailoring the FPA to the requirements of the given instrument. This is largely due to its tunable bandgap energy, and therefore cutoff wavelength. Increasing the content of Cadmium relative to Mercury shifts the cutoff wavelength towards the visible part of the spectrum. HgCdTe detector material thickness can be limited to $10\,\mu$m to $20\,\mu$m because it is a direct bandgap semiconductor which limits charge diffusion and bulk dark current generation (Norton 2002).

The FPAs developed for *SNAP/JDEM* have a cutoff wavelength of $1.7\,\mu$m though $2.0\,\mu$m and $2.5\,\mu$m detectors were developed briefly prior to the end of the R&D project. The development was initially based on and run in parallel to the work on

Wide Field Camera 3 (WFC3) detectors that were eventually installed on the *Hubble Space Telescope* (*HST*). This allowed successful innovations to be shared between the two and converged into the final WFC3 design which served as the basis of the final *SNAP* 1.7 μm cutoff devices. The goal of the R&D project was to develop low read noise (about 5 electrons), low dark current (about 2 - 3 electrons per hundred seconds) and high quantum efficiency ($> 60\%$) NIR devices. These goals have largely been met or exceeded (Schubnell et al. 2006; Brown 2007).

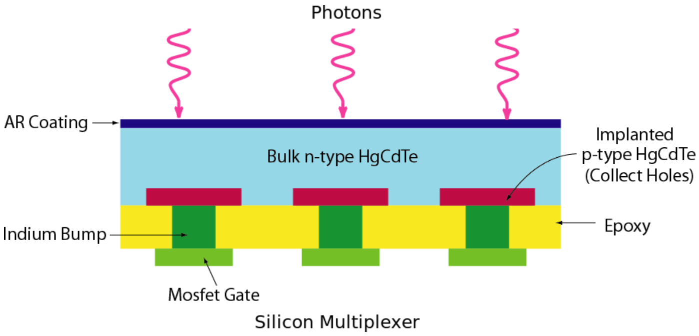


Figure 1.29: A diagram of a NIR detector pixel. Photons pass into the HgCdTe bulk where they generate electron – hole pairs. The holes are collected at the gate of a readout MOSFET electrically connected to the pixel diode with an Indium bump encased in back filled epoxy.

The *SNAP* FPAs built by TIS have 2048 × 2048 pixels with a pixel pitch of 18 μm. They are hybridized to a HgCdTe Astronomy Wide Area Infrared Imager (HAWAII)-2 readout integrated circuit (multiplexer or mux) known as H2RG. These devices are grown on a Cadmium-Zinc-Telluride substrate for good lattice matching after which the substrate is removed and an anti-reflective coating is applied. The substrate layer cuts off the QE at short wavelengths and is highly susceptible to cosmic ray–induced luminescence and hence additional dark noise (e.g. Brown 2007). Substrate removal and anti-reflective coating allow for good detector response from the UV to NIR. Figure 1.29 shows a schematic of a pixel in the H2RG line of detectors. Photoelectrons are generated in the HgCdTe bulk material and collected with an electric field at the MOSFET readout gates. Figure 1.30 shows a simplified schematic of the unit readout cell. The bulk semiconductor (represented as a diode in the figure) is biased by the substrate voltage, DSUB. The pixel is read out when it is addressed; the horizontal clock (not shown) addresses the readout channel column and the vertical clock (VCLK) selects the pixel row. The reset occurs when the "reset enable" line is asserted. Interestingly, reads can be performed simultaneously with resets. Unlike CCDs, the overall readout process is non-destructive. Charges collected at each pixel's MOSFET gate raise its source voltage which can then be read out multiple times. In principle, this charge is not lost in the process. There are subtleties involved in the device readout that can impact the image quality which can be addressed with

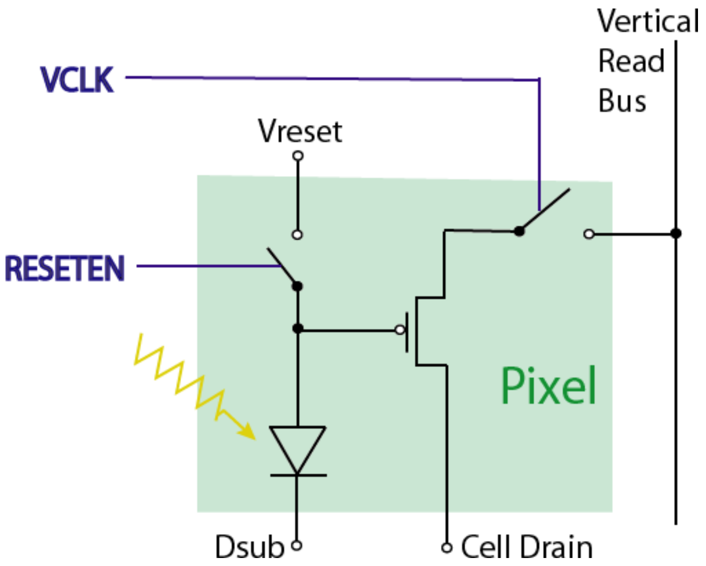


Figure 1.30: Schematic of an H2RG pixel and readout system.

proper electronic designs (e.g. Albert et al. 2008).

As the use of these FPAs increased, new effects were noticed and theorized. This prompted us to conduct additional and novel tests aimed at fully characterizing these detectors under their operating conditions. My work was focused primarily on measuring and understanding reciprocity failure and on measuring the intrapixel response in *SNAP* devices. I also developed statistical tools for detector defect characterization described in Appendix B. Reciprocity failure is discussed in great detail in Chapter 5 while detector–scale intrapixel response variation is briefly addressed in Appendix C with an early design document in Biesiadzinski et al. (2010). Also see Barron et al. (2007) for discussion of intrapixel variations in individual pixels. Below I motivate the work on these effects by describing their impact on SNe cosmology.

1.8.2 Effects On Type Ia SNe Measurements

Much of my work on reciprocity failure and intrapixel response variations was driven by calibration needs for Type Ia SNe measurements.Below I discuss some of the SNe measurement uncertainties induced by these effects in more detail. It must be noted that both effects will also play a direct role in weak lensing measurements, though their characterization is more complex and has not been studied extensively by me. In addition, reciprocity failure may have an effect on spectroscopy used for BAO by altering the continuum as well as the depth/height of absorption/emission lines.

Work has been ongoing on dealing with intrinsic systematics of SNe light curve mis-calibration, host galaxy subtraction, redshift evolution, extinction and other biases (e.g. Kim et al. 2004). Similar effort needs to be devoted to instrument calibration as it can also significantly affect the recovered cosmology. Reciprocity failure is a flux dependent non-linearity in the observed signal. For two astronomical sources where an equal number of photons has been observed, over a long exposure time (dim source) and a short exposure time (bright source), one expects to observe an equal number of collected electrons (the signal). However, due to reciprocity failure, the photons from the dim source will lead to a smaller number of electrons collected and therefore a lower signal measurement. In a nutshell, it makes dim objects appear dimmer than they really are on a device suffering from it. If not corrected for this will bias measurements of distances using standard candles like SNe since such measurements assume that source dimming is caused purely by increasing luminosity distance. An example is the measurement of H_0 (Riess et al. 2009) made with Near Infra-Red Camera and Multi-Object Spectrometer (NICMOS) which suffers from reciprocity failure (Bohlin et al. 2005). A simple test can display the cosmological implications of reciprocity failure. Figure 1.31 shows the distance modulus bias induced by a 1% and 5% reci-

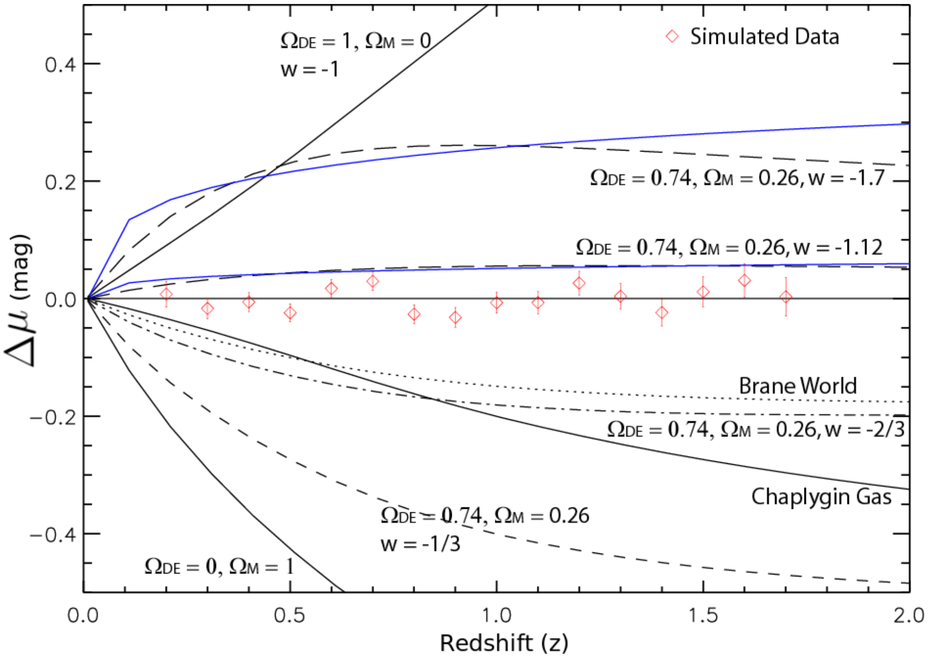


Figure 1.31: The effects of reciprocity failure on the distance modulus measurements. This figure shows the same deviations of the distance modulus from a ΛCDM universe for various cosmological models as Figure 1.26. In addition it shows the *biased* measurements for a ΛCDM universe where the bias is caused by a 1%/decade (lower blue line) and a 5%/decade (upper blue line) reciprocity failures.

procity failure models (in blue) acting on a detector in a ΛCDM universe alongside various cosmological models. This figure shows that a 1% reciprocity failure causes a ΛCDM universe to resemble a universe with the dark energy equation of state parameter w=-1.12 and a 5% reciprocity failure causes it to resemble a universe with w=-1.7. Larger reciprocity failure values (as have been observed) could potentially reproduce the behavior of an even faster accelerating universe. See chapter 5 for the discussion of our investigation of this effect.

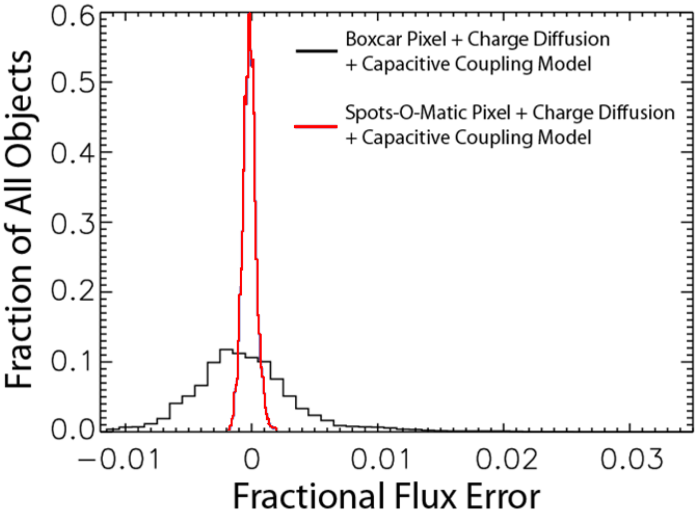


Figure 1.32: The distribution of fitted point source flux errors using predictions of Spots-O-Matic data quality. Sources with a FWHM=0.22 × pixel width were convolved with high resolution Spots-O-Matic data. Their fluxes where then recovered using either perfect pixel assumptions (black histogram) or Spots-O-Matic predictions (red histogram).

Intrapixel response variations, also referred to as subpixel structure, are important for under-sampled telescopes like *WFIRST* or even *Euclid*. Its measurement is in some respects analogous to traditional flat fielding. Some portions of a pixel may be more sensitive than others. When a PSF whose size is comparable to or smaller that the pixel size falls onto such a high sensitivity region, the recorded flux will be higher than average. If the source is imaged with multiple pixels as a part of dithering strategy, intrapixel response will increase statistical uncertainty of its flux and the uncertainty of all SNe fluxes in aggregate. Intrapixel variations have previously been measured in small, cosmetically good, sections of NIR devices using the Spot-O-Matic (Barron et al. 2007). However, that system cannot characterize an entire device in less than a year, a time period that is impractical for a survey telescope utilizing many such detectors like *SNAP*. For this reason I worked on constructing and using the Spots-O-Matic, a multi-spot projector where spots substantially smaller than the pixel pitch could measure the response inside of each pixel. This information could then be used to apply a correction to all pixels within every image. The predicted impact of Spots-O-Matic data on point source flux recovery is shown in Figure 1.32. For this test, $\approx 1\,\mu$m resolution Spot-O-Matic data was convolved with a PSF with a FWHM about 0.2 × pixel size (that is, a FWHM $= 4\,\mu$m) simulating a highly undersampled telescope. Such radical undersampling was briefly considered for *SNAP* because it would allow for a greatly increased survey area though it was subsequently rejected. The flux from the convolution of the pixel substructure with the PSF was then recovered by fitting a model using a perfect pixel assumption (black histogram) or the predicted measurements from the Spots-O-Matic (red histogram). The flux error assuming a perfect pixel is not large; most sources are well within ±1% residual. However the use of the simulated Spots-O-Matic reduces these residual by a factor of 5. The real Spots-O-Matic did not achieve this expected resolution. Nevertheless it would still reduce the residuals by at least a factor of 2 relative to perfect pixel assumptions. However, a far more important observation from the Spots-O-Matic was the great variety of pixel responses that appear to be related to the pixel readout mechanism. See Appendix C for a more complete discussion. This data would also aid with weak lensing measurements though its exact impact has not been tested.

CHAPTER 2

HaRMLSS: The SZ Simulation Framework

Clusters of galaxies, the most massive gravitationally bound structures in the universe, can serve as powerful probes of cosmology. They can be observed in a multitude of ways; as overdensities in the number of galaxies (e.g. Abell 1958; Koester et al. 2007a, and many more), X-Ray emission due to a hot intracluster medium (ICM) gas, Sunyaev–Zel'dovich (SZ) effect distortions of the Cosmic Microwave Background (CMB) (SZ– Sunyaev & Zeldovich 1972; Birkinshaw 1999; Carlstrom et al. 2002) and galaxy shape shears induced by gravitational lensing. Each of these methods has its strengths and weaknesses but their joint analysis promises to bring great advantages (e.g. Cunha 2009; Rozo et al. 2009a; Wu et al. 2010).

Due to the complicated nature of clusters of galaxies and the many ways they can be observed, it is necessary to carefully study them in simulations. Here we present a flexible and generalizable simulation framework for adding SZ observables into existing N-body simulations in order to test joint optical–SZ analysis performance. HAlo–Resolved Millimeter-wave Layered Sky Simulation (HaRMLSS) framework is designed specifically to address the possible systematics in SZ observations without making computationally costly assumptions about gas physics required for hydrodynamical treatments (e.g. Springel et al. 2001; White et al. 2002; Stanek et al. 2010; Kay et al. 2012). Instead, many different realizations of the SZ effect can be generated for a single dark matter simulation enabling us to study various effects.

The HaRMLSS simulation framework was developed specifically to aid in the joint analysis by the Dark Energy Survey (DES) (The Dark Energy Survey Collaboration 2005) and *South Pole Telescope* (*SPT*) (Carlstrom et al. 2011; Staniszewski et al. 2009) collaborations. While HaRMLSS can be used on its own (Biesiadzinski et al. 2012) or with additional simulations (Bleem et al. 2012), it is primarily meant to be used in tandem with the optical observable simulations generated for DES, namely

ADDGALS (Busha & et al. 2013; Wechsler & et al. 2013). There is also an ongoing effort aimed at generating X-ray counterparts to these simulations which will complete the set of cluster observables available to us.

Various teams have approached the problem of studying the SZ sky in a holistic fashion (e.g. Sehgal et al. 2010; Delabrouille et al. 2013). They range from full hydrodynamical simulations, through semi-analytic gas prescriptions added on top of dark matter halos (e.g. Bode et al. 2009) to random SZ profile placement. HaRMLSS paints N-body simulated halos with gas pressure profiles in order to generate the SZ signal. This approach, while simple, is very flexible because the signal is not generated from some underlying assumption of gas physics. It is instead matched to the desired inputs. Hence it is relatively easy to generate simulated skies with widely different SZ – mass scaling without having to run a separate hydrodynamic simulation. Because of that, HaRMLSS is ideally suited for systematic studies of the interplay between observational and theoretical effects that have often been lacking.

2.1 Large Scale Structure Simulations

The basis for SZ signal generation in HaRMLSS are simulated dark matter halos provided by N-body simulations. These simulations start with a matter power spectrum at a relatively high redshift that are then evolved in time assuming Newtonian gravity unless the goal is to test theories of modified gravity (e.g. Kravtsov & Borgani 2012; Dvali et al. 2000). Once a simulation has reached the present, dark matter overdensities are identified as halos (e.g. Behroozi et al. 2013) that are thought to host clusters of galaxies and intracluster gas. HaRMLSS has minimal requirements for N-body output catalogs; halo position, mass and redshift. Halo velocities are required for kSZ generation (see §2.2.2.2). Additional information like subhalo positions, halo particles' velocity dispersion, dark matter concentration and nearest neighbor separation can potentially be used in the future to improve the realism of HaRMLSS. Because of this simplicity of use, practically any halo catalog output by an N-body simulation can be used to generate millimeter-wave sky maps using HaRMLSS. For the purposes of the discussion presented here, we used the dark matter only simulations created for the DES collaboration.

As mentioned previously, one of the driving forces behind this project was the need for simulation complementary to Adding Density Determined GAlaxies to Lightcone Simulations (ADDGALS) (Busha & et al. 2013; Wechsler & et al. 2013) simulations of observed optical catalogs. This algorithm is tuned on high resolution N-body simulations to reproduce galaxy luminosity function and uses observed galaxy colors and their evolution to assign observable magnitudes to simulated galaxies. The simulated galaxy catalogs are not currently used by HaRMLSS in any way though plans are underway to include correlations between galaxy and SZ observables in future versions.

2.2 Sky Model Components

HaRMLSS works by layering various backgrounds and foregrounds together with a map of pure SZ signal created with resolved gas pressure profiles. These profiles are placed at dark matter halo locations in the underlying N-body simulation. These layers are independent though in practice the framework is split into two parts, SZ map-making followed by the addition of millimeter-wave backgrounds and foregrounds. At the second step, the instrumental effects discussed in §2.3 are also applied. Nevertheless, separate map layers can be output for diagnostic or reuse purposes with Figure 2.18 showing power spectra of such individual layers. In this section we limit the discussion to generic sky simulation that may be used by any observatory. This sky model is composed of the primary CMB, halo mass-dependent SZ signal, point sources emitting in millimeter wave and galactic foregrounds.

2.2.1 Primary Cosmic Microwave Background

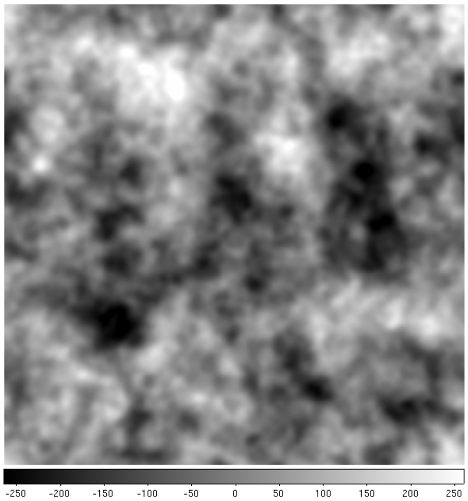


Figure 2.1: Example of a simulated CMB map. This is a $5^\circ \times 5^\circ$ simulation generated using a concordance ΛCDM cosmology inputs to CAMB. The scale ranges from -260 μK to 260 μK.

The single largest background, as well as the "backlight" making SZ observations possible, is the primary Cosmic Microwave Background (CMB). For our simulations we generate random realizations of the CMB that reproduce the power spectra obtained from Code for Anisotropies in the Microwave Background (CAMB)[1] (Lewis et al. 2000) and assume that the fluctuations are Gaussian. CAMB provides us with the lensed power spectra of the temperature anisotropies C_{TT}, the temperature and E mode polarization correlations C_{TE}, the E mode polarization anisotropies C_{EE} and B mode polarization anisotropies C_{BB}. It should be noted that this lensing is not correlated with the particles used in the N-body simulation but is instead a random realization.

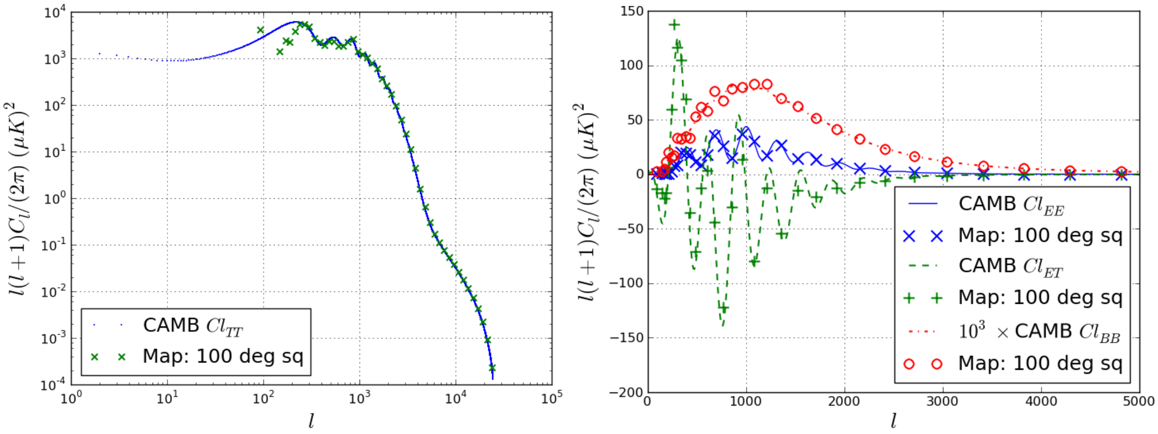


Figure 2.2: Comparison of spectra generated using a flat-sky approximation to the simulation inputs. Left panel: the temperature map spectrum from our simulation (green × symbols) along with the input C_{TT} spectrum from CAMB (blue points). Right panel: the E mode, B mode and correlated TE mode map spectra (blue × symbols, red circles and green + symbols, respectively) compared to CAMB inputs (blue solid line, red dashed-dotted line and green dashed line, respectively).

The simulated CMB maps are generated in the flat-sky approximation. The sky footprint is defined by the N-body simulation and/or by the observational program that is being simulated. This footprint is divided into square fields no larger than 10° × 10° in size. Horizontal and vertical Fourier frequency grids of map size are generated and used to compute radial Fourier frequencies. These are then mapped to the ℓ modes used by spherical harmonics modes returned by CAMB. This creates a two-dimensional CMB power spectrum in a Fourier basis instead of a spherical harmonic basis, the flat-sky approximation, though there is no m mode dependence yet. This spectrum is then multiplied by a Gaussian-random field to introduce this feature. Separate Gaussian-random fields are used for the temperature and B mode

---

[1] http://camb.info/

maps. The E mode maps use a combination of the temperature random field and an independent Gaussian-random field in order to imprint the correlation described by the C_{TE} spectrum. Finally, the E and B mode maps are rotated into Stokes Q and U maps and inverse-Fourier transformed back into angular space along with the temperature maps. An example of a $5° \times 5°$ primary CMB map is shown in Figure 2.1. All of the resulting maps match the input CAMB power spectra; an example comparing a 100 sq. degree map power spectra to the inputs is shown in Figure 2.2. The agreement between our maps and CAMB of the C_{TT} spectra in the left panel and C_{TE}, C_{EE} and C_{BB} spectra in the right panel are excellent for ℓ larger than few hundred where the map area limits the computation.

The drawback of the flat-sky approximation is that each field is generated independently. That is, the CMB fluctuations are not continuous across boundaries of adjacent fields. Future implementations of HaRMLSS may use Hierarchical Equal Area isoLatitude Pixelization (HEALPix)[2] (Górski et al. 2005) functionality to create full sky primary CMB maps.

2.2.2 Clusters

As discussed in the Introduction, §1.6.2, clusters of galaxies contain hot ICM that inverse-Compton scatters CMB photons (Sunyaev & Zeldovich 1972; Birkinshaw 1999; Carlstrom et al. 2002). Below we discuss how the SZ signatures of galaxy clusters, both thermal and kinetic, are included in our simulations.

2.2.2.1 Thermal SZ Effect

The simulation pipeline described in this paper depends on the ability to place a thermal Sunyaev–Zel'dovich (tSZ) profile at the RA, DEC and redshift of an N-body Halo. The tSZ signal is a distortion of the primary CMB temperature, $\Delta T_{tSZ}/T_{CMB}$ given by

$$\frac{\Delta T_{tSZ}}{T_{CMB}} = f(\nu) y \ . \tag{2.1}$$

$f(\nu)$ describes the frequency dependence of the SZ distortion (including relativistic corrections which we ignore in our work) and y is the Compton parameter

$$y = \frac{\sigma_T}{m_e c^2} \int n_e k_B T_e \, \mathrm{d}l = \frac{\sigma_T}{m_e c^2} \ P_e(r) \ , \tag{2.2}$$

---

[2] http://healpix.jpl.nasa.gov

where σ_T is the Thomson cross-section, m_e is the electron mass and c is the speed of light. The product of the electron number density, n_e, Boltzmann's constant, k_B, and electron temperature, T_e, is equal to the electron pressure. The integral is performed over the line-of-sight l. We will refer to this integrated pressure as $P_e(r)$ since it varies as a function of distance, r, from the cluster center.

There is in principle a lot of flexibility in choosing the exact form of the integrated pressure profile $P_e(r)$ which defines the shape of the tSZ distortion for a given halo of size R_{500} at redshift z. It can be based on theoretical predictions or observations. We chose to base our tSZ profiles on the measurements described in Arnaud et al. (2010). Bonamente et al. (2012) compares the pressure profiles of 25 massive relaxed clusters observed in X-ray and with the Sunyaev–Zel'dovich Array (SZA) and find that they agree well with the Arnaud et al. (2010) profile up to R_{500}. They found that galaxy cluster pressure profiles scaled to remove mass and redshift dependence have a roughly universal shape which can be described by a generalized NFW (GNFW) profile $\wp(\frac{r}{R_{500}})$ (Nagai et al. 2007; Arnaud et al. 2010):

$$\wp(\frac{r}{R_{500}}) = P_0 \left(\left(c_{500}\frac{r}{R_{500}} \right)^{\gamma} \left[1 + \left(c_{500}\frac{r}{R_{500}} \right)^{\alpha} \right]^{(\beta-\gamma)/\alpha} \right)^{-1} . \tag{2.3}$$

Parameter P_0 is the overall normalization of the profile and c_{500} is the profile concentration. Meanwhile parameters γ, α, β are the central ($r < \frac{R_{500}}{c_{500}}$), intermediate ($r \approx \frac{R_{500}}{c_{500}}$) and outer ($r > \frac{R_{500}}{c_{500}}$) slopes. The parameters that best fit the observed data set are given in equation 12 of Arnaud et al. (2010). That generalized profile can then be converted to pressure via equation 13 of their work. It is reproduced below:

$$P_e(r) = 1.65{\times}10^{-3} E(z)^{8/3} \left(\frac{M_{500}}{0.7 \times 3 \times 10^{14} h^{-1} M_{\odot}} \right)^{2/3+0.12} {\times}\wp(\frac{r}{R_{500}}){\times}0.7^{-2} h^{-2} keV cm^{-3} , \tag{2.4}$$

with Equation 1.39 for $E(z)$ including a parametrization of dark energy as

$$E(Z) = \sqrt{\Omega_R(1+z)^4 + \Omega_M(1+z)^3 + \Omega_k(1+z)^2 + \Omega_\lambda(1+z)^{3(1+w_0+w_a)} e^{-3\ w_a \frac{z}{1+z}}} , \tag{2.5}$$

which reduces to

$$E(Z) = \sqrt{\Omega_M(1+z)^3 + (1-\Omega_M)} , \tag{2.6}$$

in concordance ΛCDM cosmology that is usually, but not always, assumed in HaRMLSS. Combining equations 2.2, 2.3, 2.4 and 2.5 with a set of GNFW parameters allows us to create a tSZ profile scaled to an appropriate simulated halo mass, redshift and angular diameter distance. This profile is then placed in our map at that halo's location.

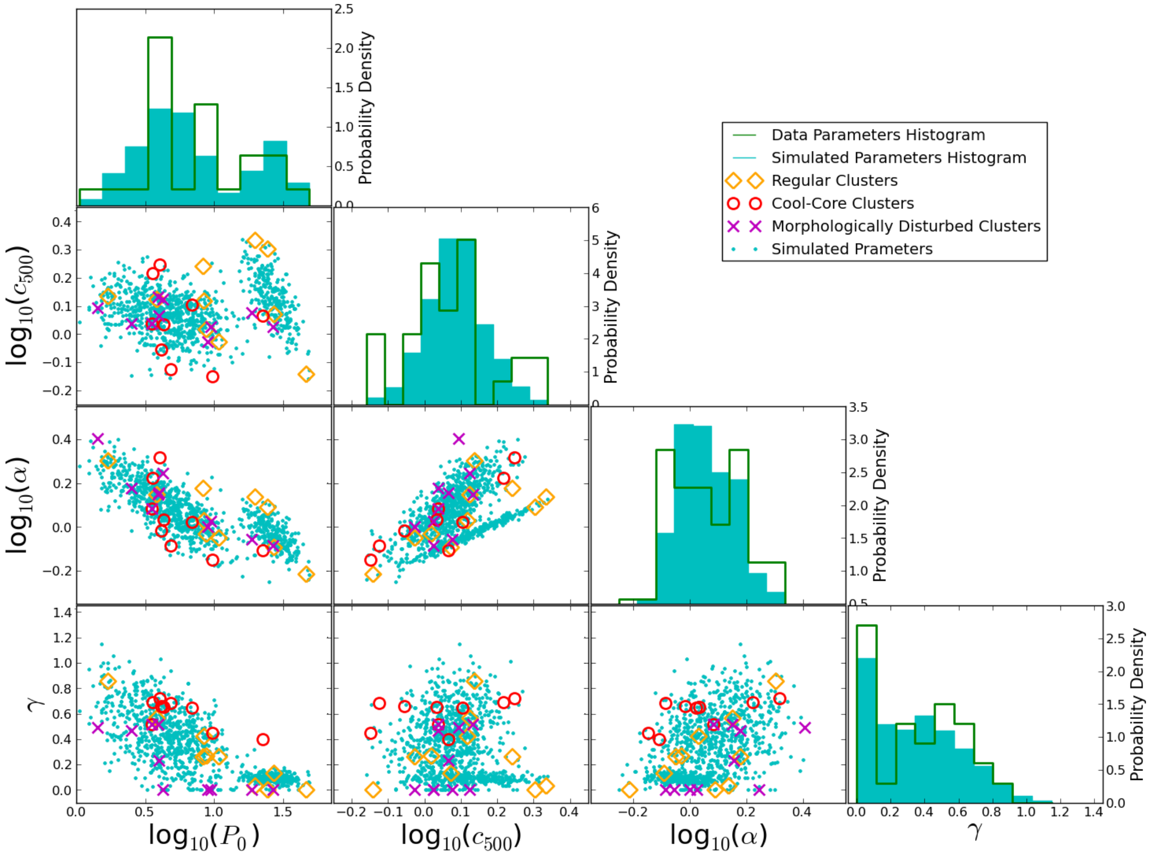


Figure 2.3: Random realizations of GNFW parameters, P_0, c_{500} α and γ (β is held fixed at 5.49) from Table C.1 of Arnaud et al. (2010) overplotted with the original parameters split into regular (orange diamonds), cool core (red circles) and morphologically disturbed (purple × symbols) populations. The random realizations are displayed as cyan dots. In addition, the histogram of each of the original parameters is drawn with a green line and the histogram of randomly drawn parameters is drawn with cyan bars.

The simplest application of our simulation pipeline takes the average GNFW parameters from equation 12 of Arnaud et al. (2010). In addition HaRMLSS supports splitting the simulated halo population into regular, cool core and morphologically disturbed objects with GNFW parameters given by Table C.2 in Arnaud et al. (2010) (see §2.5.1 for an application). Finally, we can generate a random GNFW parameter list reproducing the parameter correlations based on fitted values shown in Table C.1 of Arnaud et al. (2010). An example of such parameter generation is shown in Figure 2.3. These profile choices reproduce their SZ-mass scaling relation. In practice, to save time, we do not generate random shape realizations of low mass

($M_{500} < 2 \times 10^{13} h^{-1}\, M_{\odot}$) and simultaneously unresolved (R_{500} <1.25$'$) halos in large simulations (> 500 sq. degrees). Instead, the default Arnaud et al. (2010) profile shape is used and its overall amplitude is randomized. This means that at our default map resolution of 0.25$'$/pixel, R_{500} fits within 5 pixels. Very roughly, the full width at half maximum (FWHM) of these profiles are 0.2 × R_{500} so most of the SZ signal fits within 2 or 3 pixels hence this approximation should not cause any problems. Additional profile shapes, including those observed directly by SZ telescopes will be added in the future. In practice, computational time requirements constrain us to work with radially symmetric shapes.

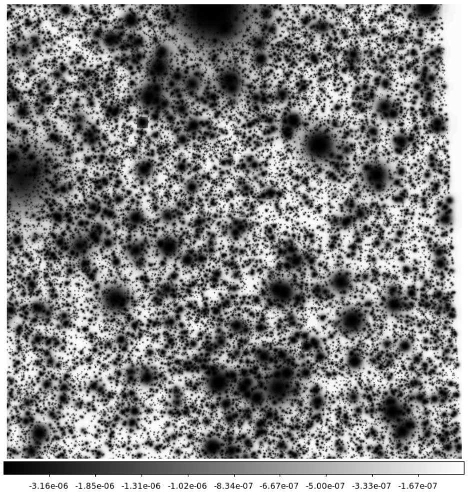


Figure 2.4: Example of a simulated tSZ $\frac{\Delta T_{tSZ}}{T_{CMB}}$ map at 150 GHz. This is a 5° × 5° simulation using the default scaling and profiles from Arnaud et al. (2010). The scale is non-linear in order to bring out the detail and ranges from -1.14×10^{-4} to 0. Note that the top right corner of the map contains no halos because none were generated there by the N-body simulation.

The current version of HaRMLSS does not lens the SZ profiles or the CMB by foreground matter and so there may exist a small mismatch between the ADDGALS galaxy cluster centers and the SZ centers and the SZ profile shapes are not sheared. The center mismatch can be addressed by using the weak lensing–induced offsets in ADDGALS cluster centers to "correct" the halo centers. The proper treatment of lensing will require additions to the framework that will be incorporated in the future. In addition, the SZ profiles are not polarized in our simulations even though such signal is expected due to the CMB quadrupole (e.g. Sazonov & Sunyaev 1999). Current and upcoming generations of CMB polarization experiments are unlikely to be able to detect this small polarization contribution.

2.2.2.2 Kinetic SZ Effect

In addition to modeling the thermal SZ effect, we also include an estimate of the kinetic Sunyaev–Zel'dovich (kSZ) effect subject to the limitations of the hydrostatic equilibrium assumption of the hot ICM gas. The kSZ signal is a distortion of the primary CMB temperature, $\Delta T_{kSZ}/T_{CMB}$, caused by the peculiar motion of galaxy clusters with respect to the primary CMB frame of reference. It is given by

$$\frac{\Delta T_{kSZ}}{T_{CMB}} = -\frac{v_{pec}}{c}\sigma_T \int n_e(l)\mathrm{d}l \ , \tag{2.7}$$

where v_{pec} is the peculiar velocity of the galaxy cluster obtained from the N-body simulation output, c is the speed of light, σ_T is the Thomson cross-section and n_e is the electron number density which must be integrated over the line-of-sight distance l. The electron number density can be obtained from the gas density ρ_g by

$$n_e = \frac{\rho_g}{\mu_e \ m_p} \ , \tag{2.8}$$

where μ_e is the mean molecular weight per free electron and m_p is the proton mass.

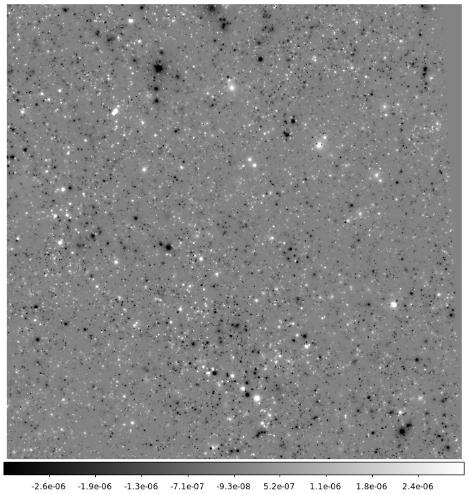


Figure 2.5: Example of a simulated kSZ $\frac{\Delta T_{kSZ}}{T_{CMB}}$ map. This is a 5° × 5° simulation. The scale is linear though clipped at low and high ends in order to bring out the detail and ranges from -3.2×10^{-6} to 3.2×10^{-6}.

The gas density at radius r can be obtained from the gradient of the generalized electron pressure profile $\mathrm{d}P_e/\mathrm{d}r$ and a Navarro–Frenk–White (NFW) (Navarro et al. 1996) dark matter density profile ρ_{NFW} by assuming that the cluster is in hydrostatic equilibrium. Under this assumption, the gas pressure gradient at radius r, related to the electron pressure gradient by the ratio of μ_e/μ (where μ is the mean molecular weight of the gas), opposes the gravitational force exerted on the gas at this radius by the combined mass of the dark matter and gas enclosed within this radius. It follows that:

$$\rho_g = -\frac{1}{4\pi G}\frac{\mu_e}{\mu}\frac{\mathrm{d}P_e}{\mathrm{d}r}r^2\left(\int_0^r(\rho_{DM}+\rho_g)r^2\mathrm{d}r\right)^{-1} \ , \tag{2.9}$$

where G is the gravitational constant and ρ_{DM} is the dark matter density. The equation is solved iteratively by first setting $\rho_{DM} = \rho_{NFW}$ and $\rho_g = 0$ inside the integral. The obtained value of ρ_g is then put back into the integral and ρ_{DM} is adjusted such that the total mass given by

$$M_{200} = 4\pi \int_0^{R_{200}} (\rho_{DM} + \rho_g) r^2 \mathrm{d}r \ , \tag{2.10}$$

does not exceed the initial cluster M_{200}. This continues until ρ_g converges such that the initial cluster M_{200} is recovered to within 0.2% when the previously computed ρ_g is inserted into equation 2.10 without further adjusting ρ_{DM}. As a result, the total density profile deviates somewhat from an NFW profile, especially for cool core clusters as seen in Figure 2.6.

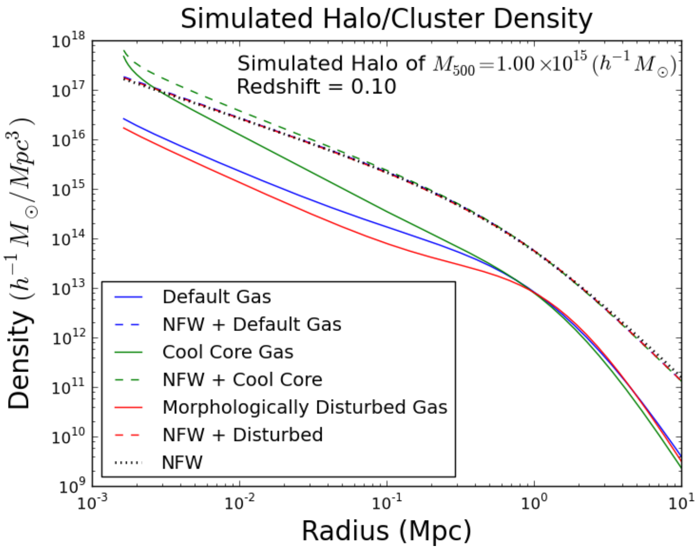


Figure 2.6: Total density profiles (dashed colored lines) compared to gas density profiles (solid color lines) after applying the iterative procedure to Equation 2.9. In addition, a pure dark matter NFW density profile (dotted black line) is shown. Note especially how the cool core total mass (dashed green) is significantly higher that the NFW density in the center where gas mass was added.

This is admittedly a fairly complicated procedure. However, it intrinsically correlates the tSZ and kSZ profiles using an intermediate dark matter profile. Future implementations of HaRMLSS may utilize the actual dark matter profiles for each halo to truly connect these signal with the underlying dark matter.

As mentioned in §2.2.2.1, we are limited by our computational resources. An approximation is therefore used to generate the tSZ and the kSZ profiles. The kSZ profiles are more sensitive to this approximation that involves pre-generating profiles. Turning a tSZ into a kSZ profile assumes a NFW dark matter profile which itself is a function of mass and redshift. It was therefore necessary to check the effect of this approximation on the kSZ profile. Figure 2.7 shows the kSZ integrated within R_{500} as a function of mass for "true" (blue circles) and approximate (green dots) kSZ profiles. In addition, the default implementation, where only low mass and unresolved profiles are approximated, is shown by the red squares. The insets show that the approximation is a factor of two low at high mass but it is not used there. At $M_{500} =$

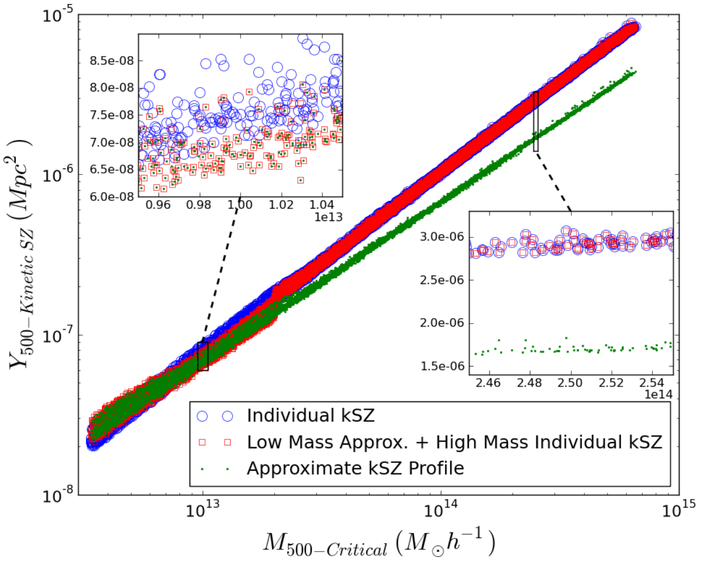


Figure 2.7: Integrated kSZ signal for profiles derived from individual (blue circles) as well as approximate (green dots) tSZ profiles. The disagreement is a factor of two at high mass. It is significantly smaller at low mass where the approximation is turned on by default (red squares).

$10^{13}h^{-1}\,M_\odot$ the error is only 7% and somewhat larger at $M_{500} = 2 \times 10^{13}h^{-1}\,M_\odot$ where the approximation first turns on. Considering that the real kSZ – mass scaling is not known and that our procedure already suffers from hydrostatic equilibrium assumptions, we do not believe that this error is significant.

2.2.3 Sunyaev–Zel'dovich - Halo Mass Scaling Relation

Of great importance to SZ observations is the scaling between the observed SZ signal and cluster mass. As described in Section 2.2.2.1 we use pressure profiles from Arnaud et al. (2010) which inherently reproduces the SZ-mass scaling relation described in that work. Our simulation pipeline is flexible enough to allow us to generate any scaling desired. Multiple relations are currently enabled and more can be added easily. They span both observational constraints (Arnaud et al. 2010; Rozo et al. 2012) with possible assumption of self-similarity. Others are based on simulations (Stanek et al. 2010; Kay et al. 2012) using various gas physics; gravity only, preheating and cooling, and feedback models. The conversion to a differing scaling relation is done by multiplying each tSZ and kSZ profile by the ratio of Y_{500} of the desired relation to Y_{500} of the default relation. Y_{500} is the SZ decrement due to a cluster integrated within an aperture of radius equal to R_{500}. In addition, different pressure profiles, when added, may also naturally lead to different SZ-mass scaling. Currently, this scaling is not well constrained observationally (e.g. Bonamente et al. 2008; Arnaud et al. 2010; Planck Collaboration et al. 2011c; Marrone et al. 2012) especially at $M_{500} < 2 \times 10^{14}h^{-1}\,M_\odot$ and hydrodynamical N-body simulations produce scaling relations strongly dependent on the gas physics prescription used (e.g. Nagai 2006; Shaw et al. 2008; Sehgal et al.

2010; Stanek et al. 2010; Battaglia et al. 2010; Kay et al. 2012).

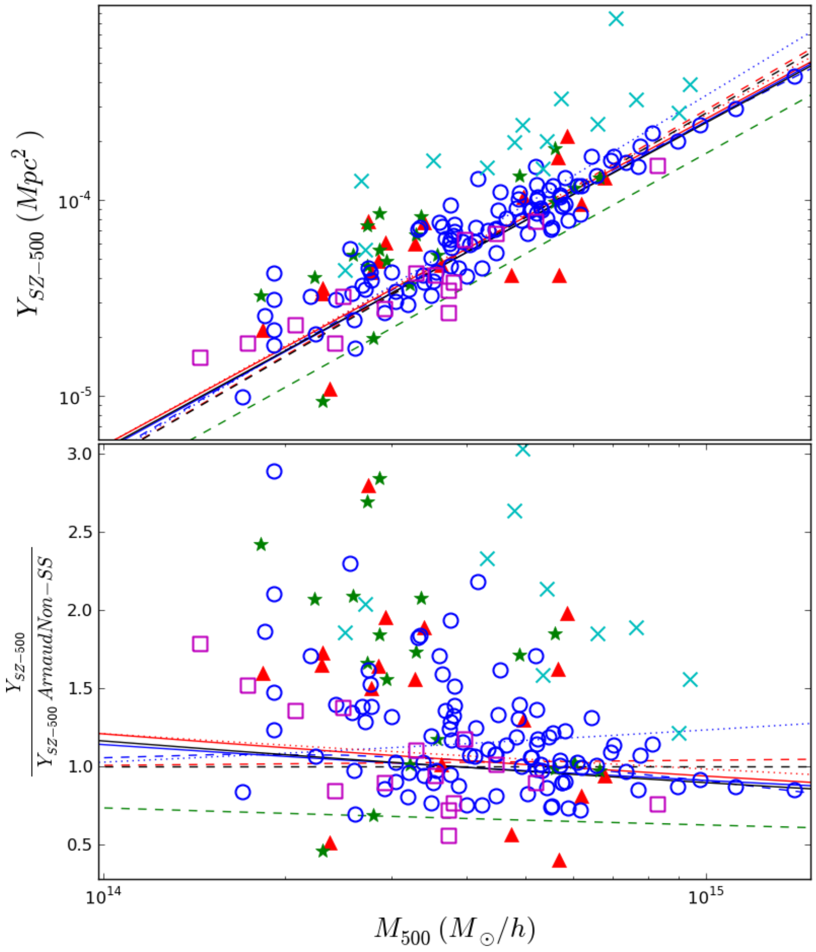


Figure 2.8: Top panel: Y_{500} vs mass for various scaling relations (lines) and data (symbols). Bottom panel: Ratio of scaling relations and data to the Arnaud et al. (2010) default scaling. Filled symbols have masses from weak lensing, open symbols have masses from X-ray measurements of Y_X and cyan × symbols are dynamical masses. See text for details on all scalings and data sets.

This is clearly illustrated in Figure 2.8 which shows multiple scaling relations and data sets in its top panel and their ratio to the Arnaud et al. (2010) default relation in the bottom panel. Scaling relations from Stanek et al. (2010) simulations are in blue: gravity only as a solid line, quadratic fit to preheating and cooling as a dashed line and linear fit to preheating and cooling as a dotted line. Relations from Kay et al. (2012) simulations are in red: gravity only as a solid line, preheating and cooling as a dashed line and feedback only as a dotted line. Relations from (Arnaud et al. 2010) are also shown, best fit as a black dashed line and self–similar - constrained fit as a solid black line. Note that while they are marked as lines, the (Arnaud et al. 2010) are based on X-ray observations and not simulations. Relation from Rozo et al. (2012) is a green dashed line and it is based on a simultaneous optimization of multiple data sets. In addition, various data sets are also plotted. Filled red triangles are weak lensing masses with SZ from CARMA (Marrone et al. 2012). Filled green stars are weak lensing masses and SZ from *Planck* (Planck Collaboration et al. 2013d). Open blue circles are Y_X masses from XMM-Newton with SZ from *Planck* (Planck Collaboration et al. 2011a,c, 2012, 2013e). Open purple squares are Y_X masses primarily from Chandra with some XMM-Newton observations and with SZ from SPT (Andersson et al. 2011). Cyan × symbols are dynamical masses and SZ from ACT (Sifón et al. 2013). There Y_{200} was provided instead of Y_{500}. Conversion was performed assuming the Arnaud et al. (2010) profile shape.

This figure must be interpreted with some caution. For instance, while the XMM-Newton X-ray follow up of *Planck* SZ clusters (blue circles in Figure 2.8) appear to follow the self-similar relation from Arnaud et al. (2010) at high masses, a lot of assumptions come into play here. The Arnaud et al. (2010) scaling itself is based on XMM-Newton observations of relaxed clusters so this apparent agreement could simply be a result of sample selection. The HaRMLSS framework's strength is its flexibility in implementing various scaling relations. This is the key to systematic checks that can be performed with any desired scaling relation and pressure profile upon its implementation.

2.2.4 Point Sources

The SZ signal can be contaminated by unresolved point sources due to galactic thermal and Active Galactic Nuclei (AGN) emissions. Following the work of Negrello et al. (2007) we split the point sources intro three categories: radio sources (e.g. De Zotti et al. 2005), protospheroidal infrared (IR) sources (e.g. Granato et al. 2004) and late-type IR sources (e.g. Gruppioni et al. 2002; Silva et al. 2004). The flux density S_ν of the sources are assumed to follow a frequency ν scaling of

$$S_\nu \propto \nu^\alpha \ , \tag{2.11}$$

where α is the spectral index. In addition, these sources exhibit correlations with galaxy clusters whose SZ signature we are simulating. Our SZ simulation framework models the above three point source populations and approximates point source – galaxy cluster correlations with simplified models. In practice, the number of sources in a narrow flux density bin is selected from a Poisson distribution and placed randomly on the map. The mean source counts are taken from densities provided by Negrello et al. (2007). An example of cumulative counts at 150 GHz is shown in Figure 2.9. Subsequently, our simple correlation models are used to correlate a small fraction of SZ point sources with galaxy clusters. It is important to note that currently the galaxy information from ADDGALS catalogs is not used to place SZ point sources. The simulated galaxy information will be used in future implementations of HaRMLSS.

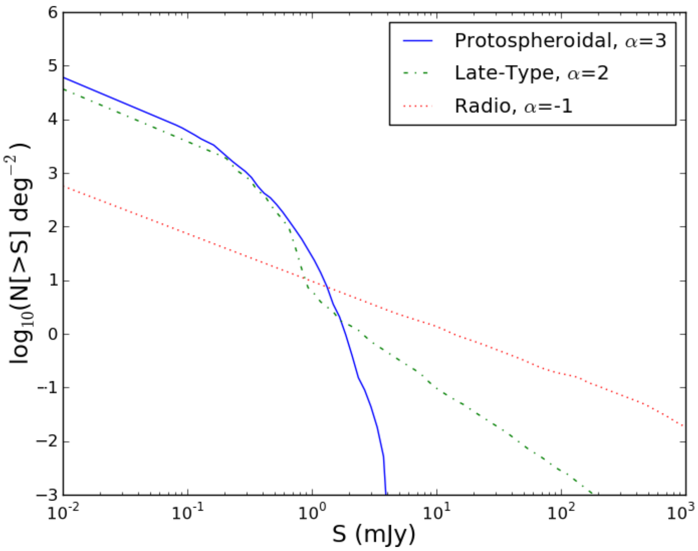


Figure 2.9: Cumulative point source counts, above radio flux S, vs flux. The blue line represents the high redshift IR protospherodial galaxies, the green dash-dotted line represents the lower redshift ($z \lesssim 1.5$) IR emitting late-type galaxies and the red dotted line represents AGN radio sources.

2.2.4.1 Radio SZ Sources

Radio sources result primarily from AGN blazars with addition of some quasars (De Zotti et al. 2005). These sources are generally associated with galaxies both in galaxy clusters as well as field galaxies. HaRMLSS uses the model from Negrello et al. (2007) which is based on De Zotti et al. (2005) to parametrize the radio source number as a function of spectral flux density. The model at 150 GHz is shown in Figure 2.9 as the red dotted line. The default implementation assumes a spectral index $\alpha = -1$. Values between -1 and 0 are consistent with observations (e.g. Vieira et al. 2010). As seen in Figure 2.10, the number and temperature of the simulated point sources agrees qualitatively with the results of Vieira et al. (2010). Note that Figure 2.10 shows point sources from all populations, not just radio, however for spectral flux densities larger than 8 mJy, radio population is the dominant one.

Correlations between radio sources and galaxy clusters are introduced via a method where a small fraction of sources are moved to coincide with halo positions on the sky. According to Vanderlinde et al. (2010) and Coble et al. (2007) clusters have about

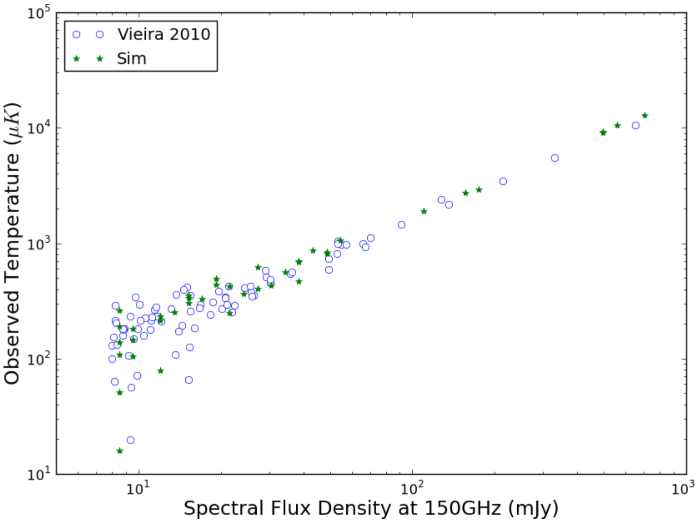


Figure 2.10: Point source temperature vs. spectral flux density of simulations and data from *SPT* at 150GHz. Data from Vieira et al. (2010). This plot demonstrates that HaRMLSS correctly converts radio flux into temperature in our simulated maps and that qualitatively, the brightest simulated object counts agree with measurements.

nine times the background probability of having a radio source within $0.5'$. For each halo more massive than $2 \times 10^{13} h^{-1}\, M_{\odot}$, an arbitrary but reasonable cutoff to prevent the exponentially rising low mass halo population from dominating, our algorithm searches for a previously generated random radio source in nine locations. If one is found, it is moved to the approximate location of the halo. This ad hoc model is not meant to make predictions about the radio source population but rather aid us in estimating radio source contamination of cluster signals.

2.2.4.2 Protospheroidal Sources

Protospheroidal SZ sources are considered to be high redshift ($z \gtrsim 1.5$), star forming galaxies undergoing an intense period of starburst (Negrello et al. 2007). These dusty galaxies absorb the light from newly formed stars and re-radiate it as infrared radiation that constitutes an SZ point source population. Following Vanderlinde et al. (2010) we chose a spectral index of 3 for this population. An example of these sources at 150 GHz is shown in Figure 2.9 as a solid blue line.

These sources also exhibit correlation with galaxy clusters due to lensing magnification. While the IR bright galaxies at high redshift are not correlated in position with galaxy clusters, they occasionally lie behind galaxy clusters along the same line-of-sight. This alignment causes the light from these galaxies to be magnified. A detailed study of this magnification is presented in Lima et al. (2010a), where it depends on the cluster mass, galaxy and cluster redshifts and their angular separation. Our framework currently treats this source of correlation in a very simplified manner. A small fraction, 0.005%, of the randomly-located protospheroidal sources that happen to lie within $0.25'$ (the resolution of our maps) of halos more massive than

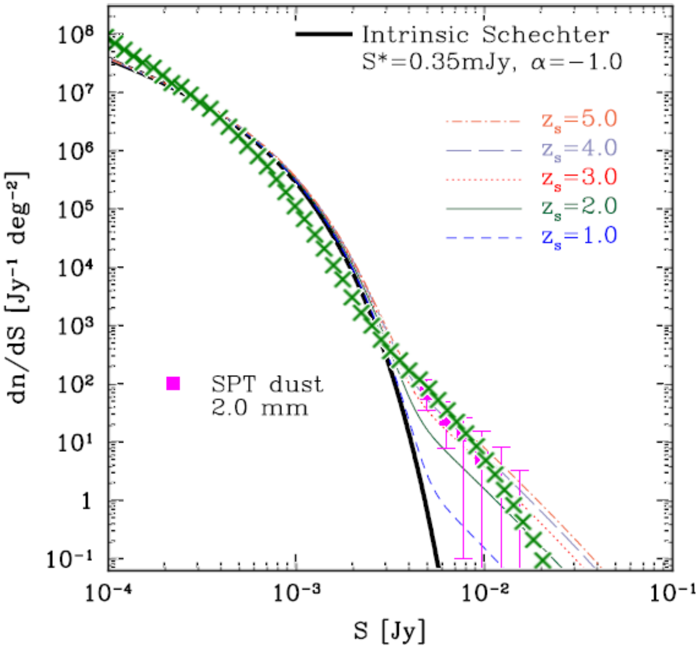


Figure 2.11: Comparison of HaRMLSS modeled lensed point source distribution (green × symbols) to data (magenta filled squares) (Lima et al. 2010b). Some of their models are also included as various colored lines.

$5 \times 10^{13} h^{-1} M_\odot$ are lensed by a factor of 10 to *roughly* reproduce the lensed source density shown in Lima et al. (2010b). This source density is reproduced in Figure 2.11 (filled magenta squares) along with their predictions (various colored lines) and our model (green × symbols). The factor of 10 was chosen as a representative magnification from Lima et al. (2010a). While far from complete, this model does provide an estimate of the impact of protospheroidal point source correlation.

2.2.4.3 Late-Type Sources

In addition to the protospheroidal source population discussed above, late-type and irregular galaxies at lower redshifts ($z \lesssim 1.5$) also occasionally enter periods of starburst (e.g. Gruppioni et al. 2002; Silva et al. 2004; Negrello et al. 2007). The density of these sources is only significant for relatively dim objects as can be seen in Figure 2.9 where they are represented by a green dot-dashed line with a spectral index of 2 (Vanderlinde et al. 2010). These sources exhibit anti-correlation with galaxy clusters in that high mass clusters have a lower star formation rate than field galaxies (Hashimoto et al. 1998; Bai et al. 2007). Since these sources do not contribute much SZ flux, we do not model this correlation and instead treat them as purely random.

2.2.4.4 Results of the Correlation Model

Figure 2.12 shows the results of our simplified SZ point source – galaxy cluster correlation model. It compares the amount of the central SZ decrement y_0 filled in within halos by uncorrelated (red histogram) and correlated (blue histogram) point sources at 150 GHz (top panel) and 220 GHz (bottom panel). The effect of correlations is not significant in agreement with conclusions drawn in Vanderlinde et al. (2010). Note that the effects of correlations are less significant at 220 GHz. This occurs because radio point source contamination is the dominant effect, and due its spectral index of -1, in contributes less at higher frequencies. Radio source contamination will be more severe at 95 GHz.

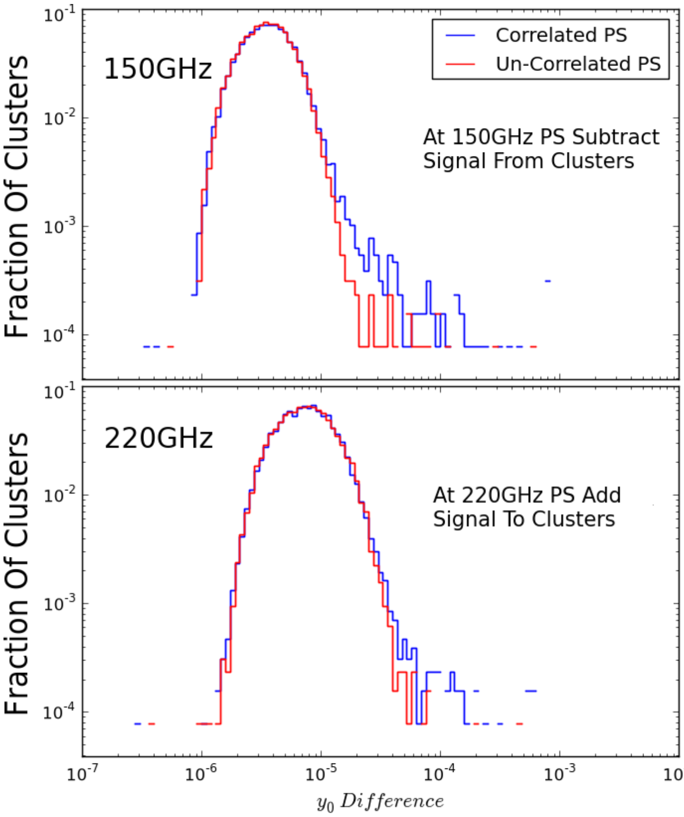


Figure 2.12: Effect of the HaRMLSS point source correlation model. We plot the fraction of clusters that had their central SZ decrement y_0 changed by a nearby point source in bins of the magnitude of y_0 change. The top panel shows results at 150 GHz and the bottom panel shows results at 220 GHz. The blue histogram displays the results including our correlation model while the red histogram shows the results without point source – cluster correlations.

2.2.5 Galactic Foreground

Galactic foreground emission results from thermal dust radiation and synchrotron emission due to the galactic magnetic field. The synchrotron is not expected to contribute much in terms of overall or polarized radiation at frequencies above 90 GHz (O'Dea et al. 2012). We therefore limited our work to the thermal emission due to dust and its polarization. O'Dea et al. (2012) modeled dust emission and polarization at 90 GHz, 150 GHz, 220 GHz, 250 GHz and 280 GHz and kindly made these models available[3]. An example temperature map at 150 GHz is shown in Figure 2.13.

---

[3] http://www3.imperial.ac.uk/people/c.contaldi/fgpol

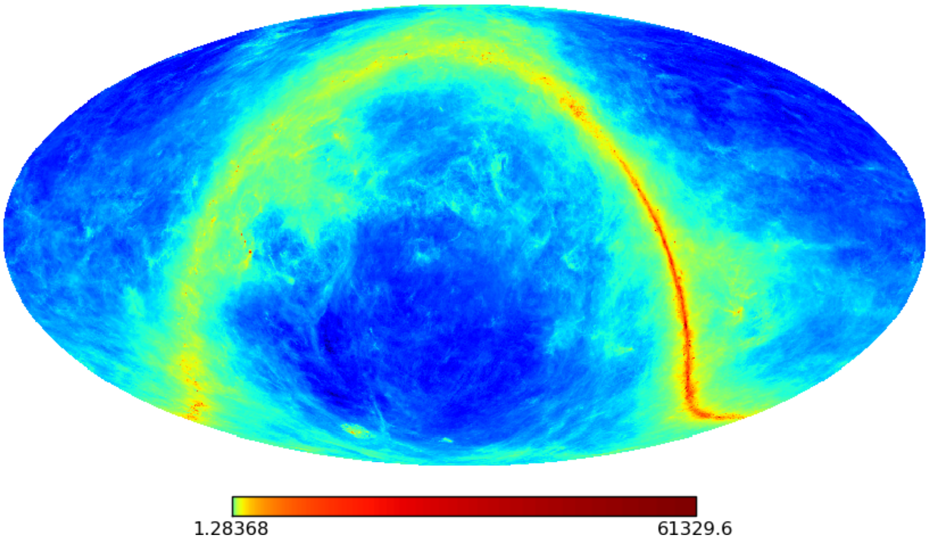


Figure 2.13: Galactic dust temperature map model in equatorial coordinates at 150 GHz from (O'Dea et al. 2012). The color scale is in μK and is not linear in order to bring out more detail.

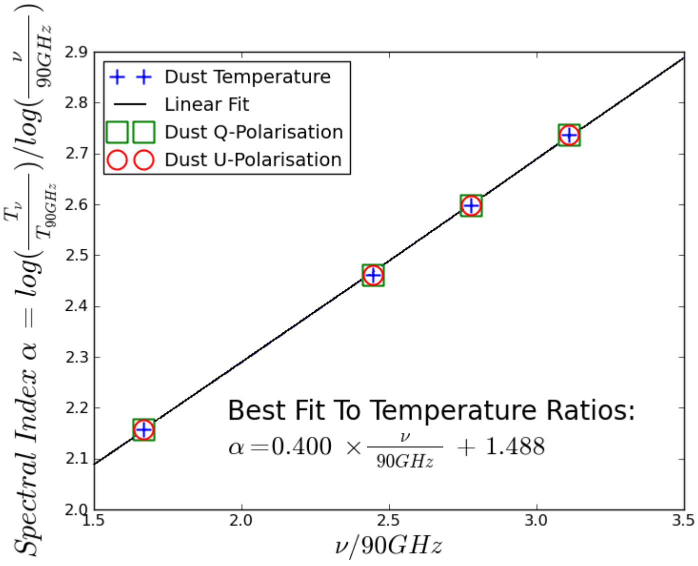


Figure 2.14: Spectral index vs frequency for average galactic dust model maps. Frequency is normalized to 90 GHz and the spectral index is computed using average temperature or polarization ratios to their value at 90 GHz. Plotted are average temperature (blue + symbols), average Q mode polarization (green squares) and average U mode polarization (red circles). Also shown is a linear fit to the temperature spectral index (solid black line).

This frequency coverage does not fully match *SPT* or *Planck* observable bands hence it was necessary to extend them to other frequencies so that the galactic foreground can be interpolated for various CMB observatories. The initial assumption that these maps, temperature as well as Q and U mode polarization, would follow the spectral dependence of Equation 2.11 proved incorrect, however a trend was observed. We found that the spectral index varies with frequency as can be seen in Figure 2.14 which shows that the spectral index α for the average maps is a linear function of frequency. The model used to derive a universal temperature parametrization was therefore

$$S_{T,\nu} = S_{T,0}\ \nu^{\alpha_0+\alpha_1\nu}\ , \tag{2.12}$$

which was converted to the log form

$$log\,(S_{T,\nu}) = log\,(S_{T,0}) + \alpha_0\ log\,(\nu) + \alpha_1\ \nu\ log\,(\nu)\ \ . \tag{2.13}$$

This linear form was solved for the temperature intensity normalization $S_{T,0}$ and the slope α_1 and intercept α_0 of the spectral index for *individual* pixels in the temperature maps. Using individual pixels instead of map averages as shown in Figure 2.14 significantly increased the model accuracy. These parameters could not be computed for the polarization maps since polarization values could be negative. Instead, the spectral index parameters for the temperature were assumed and only the Q and U polarization intensity normalizations ($S_{Q,0}$ and $S_{U,0}$ replacing $S_{T,0}$) were computed using Equation 2.12. Nevertheless, the temperature and polarization models are accurate to 1% or better when compared to the input maps. Our parametrization also allows for extrapolation of the dust model however, as with any extrapolation, one must proceed carefully.

2.2.6 Combining Model Layers

The previous sections describe the components used to simulate the SZ sky in a generic fashion that can be "observed" using specifications of existing or planned observatories. But first, these components must be put together at the correct frequency band. The first layer of the simulated observation consists of the combined tSZ and kSZ maps. The tSZ map is multiplied by the appropriate frequency factor $f(x)$.

$$f(x) = \left(x\frac{e^x+1}{e^x-1} - 4\right)(1+\delta_{SZE}(x,T_e))\ \ , \tag{2.14}$$

where x is defined as

$$x = \frac{h\ \nu}{k_B\ T_{CMB}}\ , \tag{2.15}$$

h is the Planck constant, ν is the observational frequency, k_B is the Boltzmann constant and T_{CMB} is the mean temperature of the primary CMB. In addition, Equation 2.14 includes a relativistic correction δ_{SZE} which is ignored in HaRMLSS. It's contribution is only a few percent for the most massive clusters (Carlstrom et al. 2002; Vikhlinin et al. 2009) with $M_{500} \gtrsim 1 \times 10^{15}\, M_\odot$. After this frequency scaling is applied, the SZ maps are added to the primary CMB map. Realizations of the CMB are independent of the frequency band for the same sky patch which is enforced in HaRMLSS by using the same random number seed for its generation regardless of the frequency. Next, a realization of the point source distribution is created and added to the simulated map at the first frequency "observed." Generally, the point source realization data file is saved after this first frequency is generated so that it can be reused at other frequencies. The fluxes of these point sources are scaled to the appropriate frequency according to Equation 2.11 and spectral indecies defined in §2.2.4.1, §2.2.4.2 and §2.2.4.3. If correlations between point sources and galaxy clusters are desired, halo catalogs are also provided to the framework so that it can introduce these. Following the point sources, the dust foreground described in §2.2.5 can be scaled to the appropriate frequency and added to our simulated maps. Unlike the primary CMB and point sources, this foreground is not randomized but rather fixed. In principle, the power spectrum of the galactic foreground could be used to generate its random realization. However, unlike the CMB which is constructed using a radial power spectrum, this foreground has to reproduce the two dimensional structure of the galaxy since survey strategies of observatories are explicitly designed to deal with its structure. This is not currently attempted but may be in the future. After SZ, CMB, point sources and galactic foreground are combined, at proper observational frequency, the sky model is constructed. We can now "observe" it with different telescopes. In addition, of interest are maps where SZ signal is not included. These *noise only* maps can be used to study false detection rates in simulations as well as compare the subtle effects due to the presence of clusters by using identical background, foreground and noise models.

2.3 Simulated *SPT* Observations

Below we discuss simulations of *SPT* observations in detail. The framework has also been utilized to simulate *Planck* satellite observations based on the ∼10 month results (Planck HFI Core Team et al. 2011) and the *Atacama Cosmology Telescope* (*ACT*) (see Swetz et al. (2011) for overview) observations. However, the bulk of our studies concerned the *SPT* due to its significant overlap with the DES footprint in order to prepare for joint *SPT* – DES science analysis. For this reason we have taken great care to reproduce the performance of *SPT*.

2.3.1 Instrument Beam

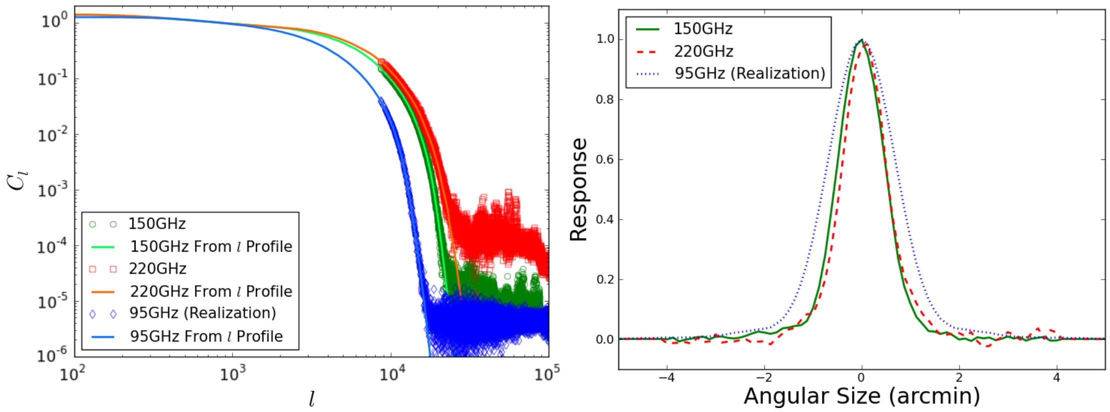


Figure 2.15: ℓ space and real space beam profiles of the *SPT*. Left panel: ℓ space profiles provided by *SPT* are shown as solid lines, light green for 150 GHz, light red for 220 GHz and light blue for 95 GHz. Also plotted are our computed profiles, green circles at 150 GHz and red squares at 220 GHz. The 95 GHz profile (blue diamonds) is computed from a realization of the ℓ space spectrum. Right panel: cross-section of the beams in real space at 95 GHz (dotted blue line), 150 GHz (solid green line) and 220 GHz (dashed red line).

The first step in simulating an instrument after the sky maps (see §2.2.6) involves convolving it with the instrumental beam in order to reproduce the telescopes resolution. HaRMLSS can generate a Gaussian beam of a specified size internally or can accept an input file with the measured instrument beam. The *SPT* collaboration has provided beam images for the 150 GHz and 220 GHz bands in Schaffer et al. (2011)[4] along with their average power spectra (in multipole ℓ). These spectra (green line

---

[4] https://pole.uchicago.edu/public/data/maps/ra5h30dec-55/

for 150 GHz and red line for 220 GHz) are compared with the spectra computed from the images at 150 GHz (green circles) and 220 GHz (red squares) in the left panel of Figure 2.15. They match as expected.

The *SPT* collaboration also provided us with the average ℓ–space spectrum of the 95 GHz beam (blue line in the left panel of Figure 2.15) but not with a two dimensional image of the beam. We were therefore forced to generate a realization of the beam at this band with a small contribution of white noise. Its spectrum is also shown in the left panel of Figure 2.15 as blue diamonds. However, unlike the other bands, the beam at 95 GHz is intrinsically radially symmetric. Still, the effect on the simulations should be minimal. The real space beams are shown in the right panel of Figure 2.15. Their size increases with decreasing frequency due to diffraction.

2.3.2 Instrument Noise

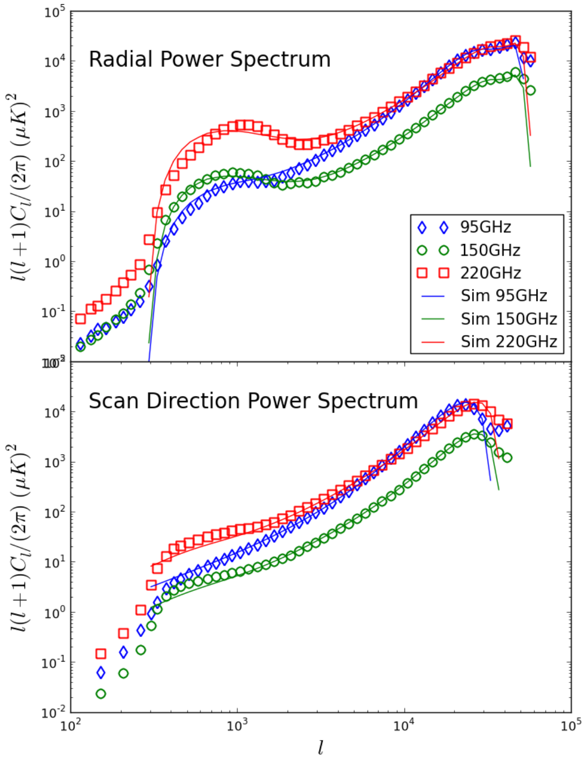


Figure 2.16: Noise PSDs of *SPT* at the three observational bands along with our parametrized models. Data at 95 GHz (blue diamonds), 150 GHz (green circles) and 220 GHz (red squares) is shown along with our model at the same frequencies. Top panel shows the isotropic power spectra. Bottom panel shows the power spectra in the scan direction of the telescope. Note that these plots also show contributions from the *SPT light* filter.

The next step of our simulation involves dressing it with instrumental noise appropriate for the telescope and frequency band. Due to the scan strategy of the *SPT*, sources of noise can contribute isotropically or one dimensionally in the direction of the telescope scans. Our parametrization of noise includes isotropic white noise, isotropic atmospheric noise with $1/\ell^3$ spectrum, scan–direction $1/f$ noise [5] with $1/\ell$ spectrum and scan–direction time–constant deconvolution noise with a spectrum proportional to ℓ^2. Instruments like *Planck* or *ACT* exhibit isotropic $1/f$ and time–constant deconvolution noises.

The exact noise level varies between the three observing bands of *SPT*. Using the *SPT* power spectral density (PSD) maps obtained from Schaffer et al. (2011) and internal *SPT* collaboration communication we have been able to approximate the telescope performance at all three frequencies. Figure 2.16 shows the actual *SPT* noise spectra at 95 GHz (blue diamonds), 150 GHz (green circles) and 220 GHz (red squares) along with our simulations shown as solid lines with the same color scheme as the data. The model parametrization is good, especially in the region of interest between ℓ of 1000 and ℓ of 10000 where the SZ signal can be extracted.

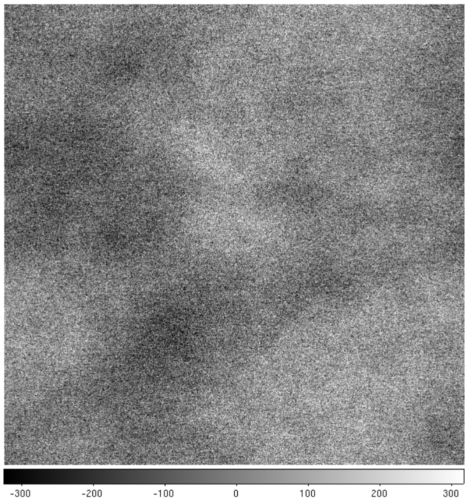


Figure 2.17: Example of a simulated *SPT* noise map at 150 GHz. This is a 5° × 5° simulation. The scale ranges from -320 μK to 320 μK. Easily visible large scale blotches are due to the atmospheric noise.

A final contribution to the noise power spectrum is the *light* filtering performed by *SPT* during data acquisition aimed at suppressing atmospheric, $1/f$ and time–constant deconvolution noise. Figure 2.17 shows a sample noise map at 150 GHz resulting from our *SPT* noise parametrization. The data is processed with a pass–band filter in the scan direction between $\ell = 300$ and $\ell = 42000$ and a high pass filter in the isotropic direction with a cut on at $\ell = 300$. The transfer functions of these filters have some curvature built in to make for smoother filtering. The filters are

---

[5] Since I am working in angular space this is actually $1/\ell$ noise. However it is generally known as $1/f$ and I shall keep that convention here.

shown as a dotted cyan line in Figure 2.18. Note that filters are applied after point source subtraction so as not to introduce ringing around said sources.

Overall, Figure 2.18 shows the power spectra of various components of our simulated $SPT-$like maps at 150 GHz. These components are astrophysical (SZ, primary CMB and point sources) and instrumental (beam impact, instrument and atmospheric noise as well as the impact of point source masking). It is interesting to observe how the point source contributions (solid orange line) are reduced first by beam convolution (dashed orange line) which acts as a low pass filter and then by real space point source masking which removes the point source power from the full simulation (black × symbols going to the black line). Galactic foreground, the newest addition to the simulation framework, is not show here as it is not significant overall.

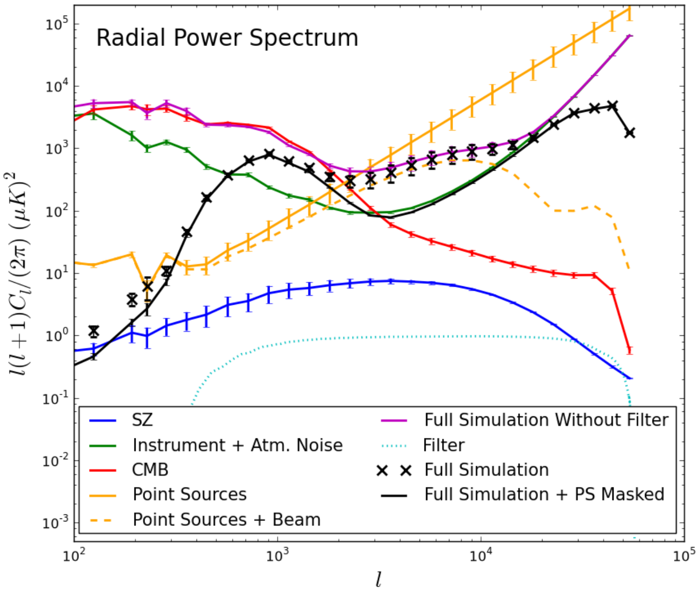


Figure 2.18: Power spectra of various astrophysical as well as instrumental contributions to simulated $SPT-$like maps at 150 GHz. Maps were output during the simulation procedure for diagnostic purposes. The pure SZ contribution is shown as a solid blue line, instrument and atmospheric noise is shown as a green line, the primary CMB is shown as a red line, and point sources are shown as an orange line. In addition, the beam-convolved point source power spectrum is also shown as a dashed orange line and the light filter is shown as a dotted blue line. The complete simulation is comprised of black × symbols and its un-filtered version is shown as a purple line. Finally, the black solid line shows the effectiveness of point source masking in the final simulation.

2.4 Model Verification

The maps generated using HaRMLSS are verified in a multitude of ways. Individual component verifications have been shown in the previous section. Here we focus on overall map checks. A sample map at 150 GHz is shown in Figure 2.19 It is scaled to show its most important features so the point sources are saturated.

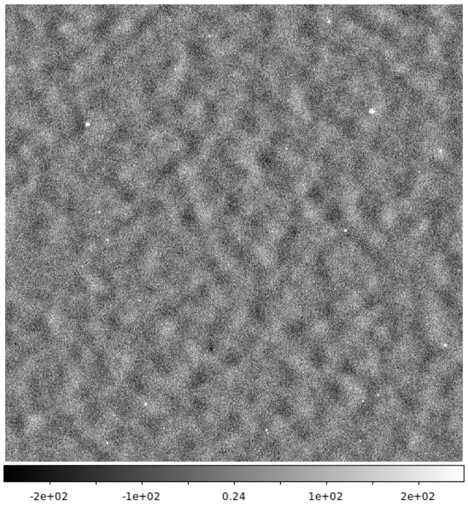


Figure 2.19: Example of a simulated *SPT* map at 150 GHz. This is a 5° × 5° simulation. The scale ranges from -250 μK to 250 μK.

2.4.1 Power Spectra Comparison To Data

We compare the power spectrum of our simulated maps (solid lines) to the Schaffer et al. (2011) data release (× symbols) in Figure 2.20. The top panel of that figure shows radial power spectra while the bottom panel shows the scan direction power spectra after point source masking. Data and simulation at 150 GHz are shown in blue and at 220 GHz are shown in green. Data at 95 GHz was not provided hence the comparison in Figure 2.16 must suffice. For completeness we also show the simulations and data prior to point source masking as dashed lines and + symbols, respectively. These simulations have larger power around $\ell = 4000$ due to one or two bright point sources. Due to the Poisson statistics of such a small number of point sources we do not consider this disagreement significant. Either way, point source masking brings the data and simulations into good agreement, especially between $\ell = 600$ and $\ell = 20000$ which encompasses the region of interest for cluster studies.

2.4.2 Y_{500} – Mass Scaling

Another check performed as part of our validation process is to make sure that the input raw SZ Y_{500} signal matches the recovered signal. We do so for all of our simu-

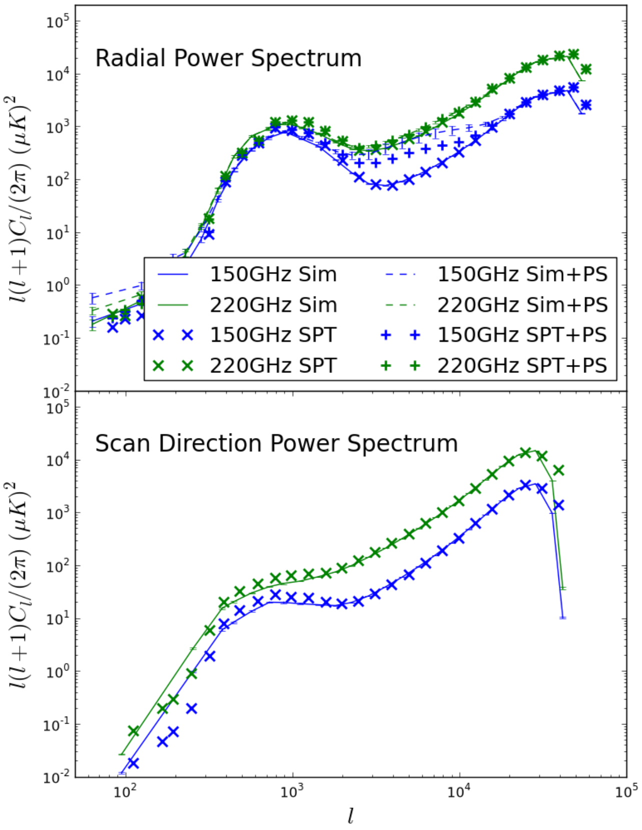


Figure 2.20: Comparison of simulated *SPT* – like map power spectra to observations by *SPT*. Top panel shows the radial power spectra while the bottom one shows the scan direction spectra. Solid and dashed lines show simulations, and × and + symbols show data, with and without point source masking, respectively. The blue symbols and lines apply to 150 GHz and the green ones apply to 220 GHz.

lations using the underlying N-body halos. This is shown in the left panel of Figure 2.21 where the recovered cylindrical Y_{500} (blue circles) vs the input scaling relation (red line) is plotted on the top. It also shows statistical and Gaussian averages as green stars and black × symbols, respectively. At the bottom, the fractional residual of the recovered integrated Y_{500} signal from the input model is shown. This is a useful tool to detect any obvious issues in profile creation. Due to the enormous influence of projection effects at low mass it is very difficult to truly judge the fidelity of our profile creation. For this reason HaRMLSS can output diagnostic simulations as shown in the right panel of Figure 2.21. For these maps we do not use the N-body halos. Instead, SZ profiles are spaced uniformly in redshift slices that eliminate all projection effects. The top plot there shows the recovered signal along with the input profiles and means in the same fashion as the left panel. But in this case, no projections are seen and the input scatter of 20% in Y_{500} is reproduced well in the fractional residual plot on the bottom.

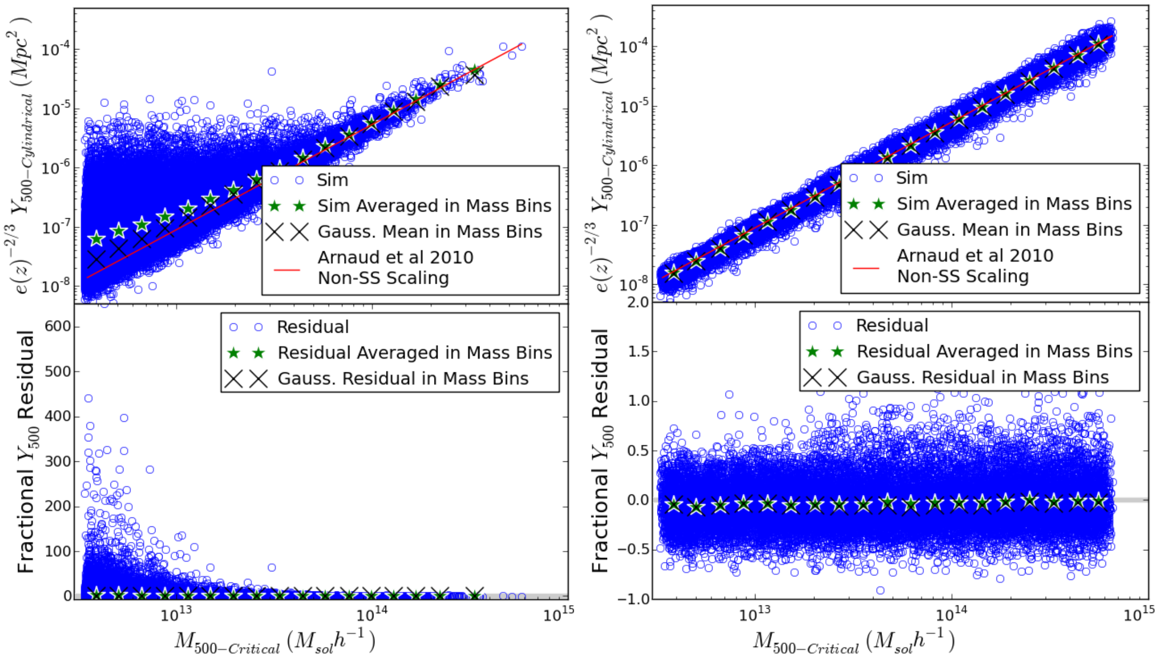


Figure 2.21: Comparing Y_{500} recovered from simulated SZ maps to the input scaling relation. Left panel shows the signal in a normal map based on N-body halos while the right panel shows the signal in a diagnostic map with well spaced profiles. The top plots show the recovered Y_{500} (blue circles), input scaling relation (red line) and statistical and Gaussian averages (green stars and black × symbols, respectively). Bottom panels show the fractional residual of the recovered Y_{500} and its means from the input scaling relation.

2.4.3 Cluster-finding

For the final validation the maps are processed with a Matched Filter (MF) (e.g., Herranz et al. 2002; Melin et al. 2006) in order to locate simulated clusters, determine the output catalog completeness and purity and determine false detection rates. We can then compare these to other simulations. Two 2500 sq. degree simulations were created with identical background, foreground and noise models. Only one of them included the SZ signal. We used the one without the SZ contributions to check the false detection rate. We plot the density of detections, in counts per degree squared vs the detection threshold (signal to noise (S/N) amplitude) in Figure 2.22. At a S/N of 5, the false detection rate is approximately 1 per 100 sq. degrees for the MF run only at 150 GHz (blue line) and slightly better for the MF run on all three frequencies (red line). This is consistent with Vanderlinde et al. (2010) and Reichardt et al. (2013) who also find that the catalog obtained using all three bands had fewer false detections just as seen in Figure 2.22.

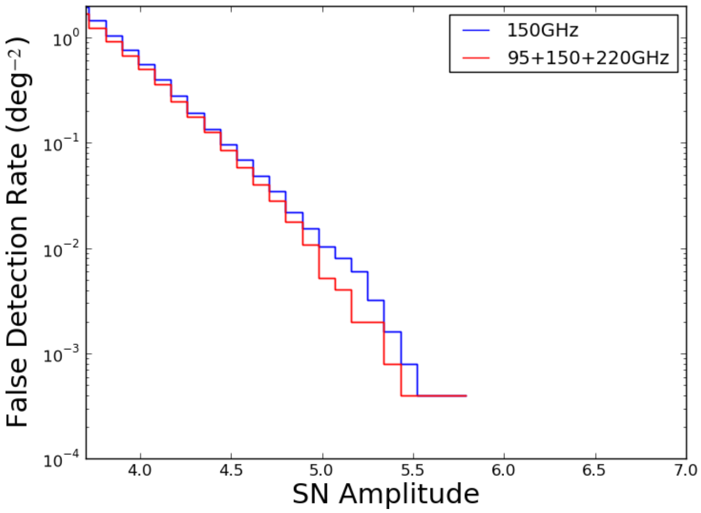


Figure 2.22: False detection rate in SZ-free simulations. The number of false detections drops as a function of detection threshold for both the single frequency (150 GHz) cluster-finder (blue line) and the multi-frequency cluster-finder (red line).

We then use the full simulations, including SZ, to determine the purity (left panel of Figure 2.23) and two dimensional completeness (right panel of Figure 2.23) of a MF run on simulations created by HaRMLSS. Purity is implicitly a function of halo mass since if there are enough low mass halos then the chance that one of them will be randomly associated with a false detection of an SZ cluster will tend towards unity. The left panel of Figure 2.23 therefore shows the purity at two cutoff masses for halo associations. The higher mass, $10^{14}h^{-1}\,M_\odot$ is more appropriate since there are few

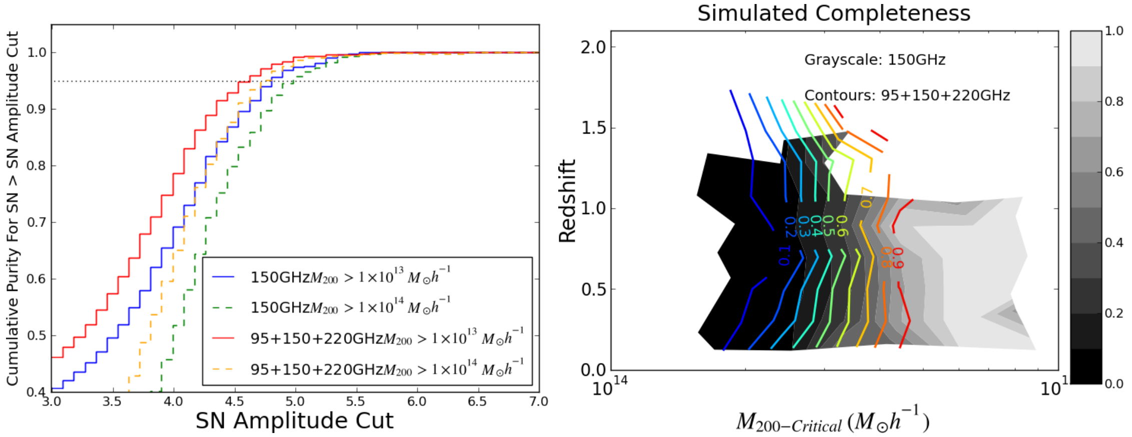


Figure 2.23: Completeness and purity of single and multi-frequency MF runs on HaRMLSS simulations. Left panel: purity as a function of detection threshold S/N. Solid blue and red lines show purities using halo mass cutoff of $10^{13}h^{-1}\,M_\odot$ for single and multi-frequency MF respectively. Dashed green and orange lines show purities using halo mass cutoff of $10^{14}h^{-1}\,M_\odot$ for single and multi-frequency MF respectively. Right panel: two dimensional completeness for single (gray scale) and multi-frequency MF. Only halos with $M_{200} \geq 10^{14}h^{-1}\,M_\odot$ were used in matching.

enough of these halos. Therefore the dashed green and orange lines show the purity vs detection threshold for single and multi-frequency MF runs respectively. The 150 GHz single frequency MF output reaches 95% purity at a cut of S/N=5 which is consistent with Vanderlinde et al. (2010). The multi-frequency MF output does a little better. The right panel of Figure 2.23 shows the two dimensional completeness in our simulations. Again, the multi-frequency MF (color contours) does somewhat better than the single frequency MF (gray scale). This completeness is considerably worse than in Vanderlinde et al. (2010). This is due completely due to their modeling of the SZ signal. Their signal amplitude is much higher, especially at higher redshift, making it much easier to find clusters there. Our scaling relations are better aligned with current observations so the completeness in Figure 2.23 should be close to the truth with the important caveat that there are essentially no constraints on the scaling relations above redshift of unity and assumptions about self-similar evolution of SZ must be made.

2.5 Sample Applications

In this section I will demonstrate how HaRMLSS can be used to study various systematic effects involving joint SZ – optical cosmological studies. Specifically, I will look at what can be learned about the impact of gas pressure profile shape variations and differences in SZ – Mass scaling relations.

2.5.1 Pressure Profile Variation

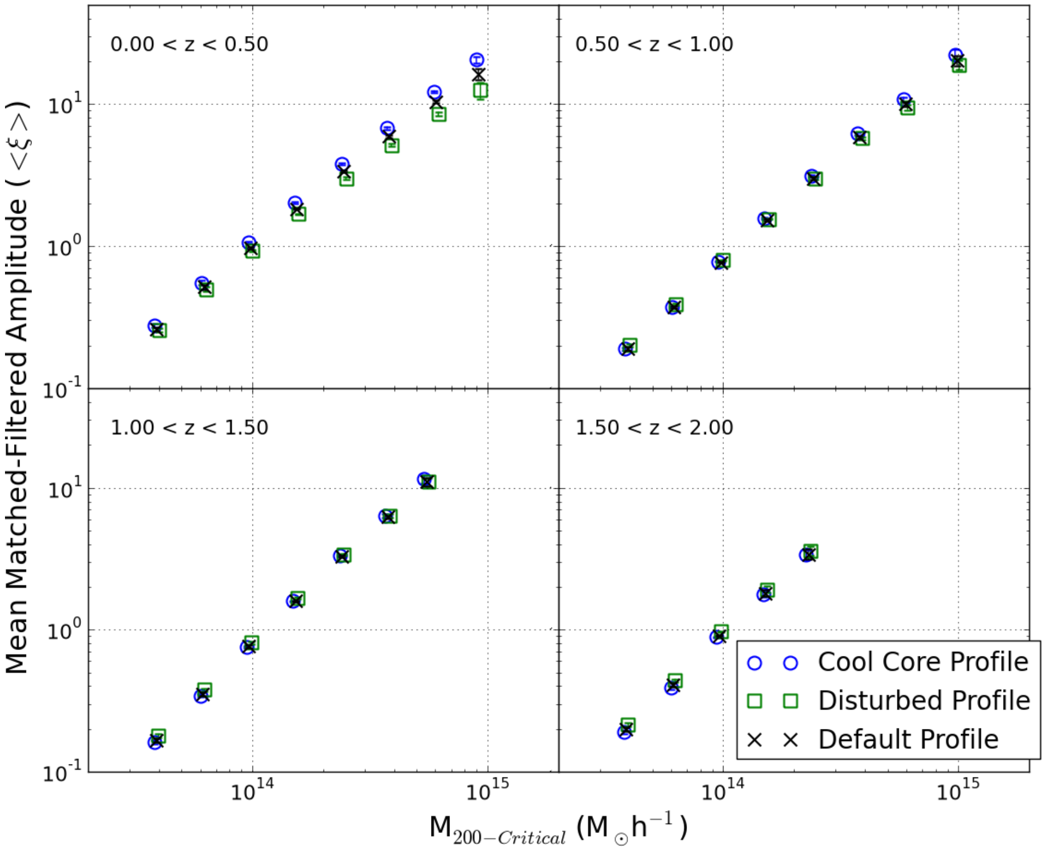


Figure 2.24: Mean MF amplitudes vs mass of stacked halos in 2500 sq. degree *SPT*–like simulations where the SZ profile shape was varied. The four panels represent the four redshift bins that our sample was divided into. Cool core clusters are represented by blue circles, morphologically disturbed clusters are represented by green squares and the default profile shapes are marked by black × symbols. Note that the low mass limit of $3 \times 10^{13} h^{-1} M_\odot$ was set by the profile shape approximation scheme discussed in §2.2.2.1.

We applied a MF to our simulated *SPT* observable maps in a similar fashion to some of the analysis performed by the *SPT* collaboration (e.g. Vanderlinde et al.

2010). Specifically, we simulated three SPT – like maps on a sky patch of approximately 2500 sq. degrees with identical underlying halo catalogs, primary CMB, point sources and instrument noise. The only difference between the three was the choice of the gas pressure profile shape. One simulation used the default shape, one used the cool-core shape and one used the morphologically disturbed shape (Arnaud et al. 2010). After these maps were matched filtered, we binned them in mass M_{200} and redshift z. The resulting mean MF amplitude is plotted vs mass in four redshift bins (one per each panel) in Figure 2.24. One can see that the data points of the three profile shapes are slightly different although the details are not clear. As an aside, it should be noted that the approximate detection threshold for individual identified clusters in SPT is $5\,\xi$ and therefore the stacked clusters reach a far lower mass limit than SPT – discovered clusters.

We wished to investigate the differences further by plotting the ratios of the MF amplitudes of the cool core and disturbed clusters to the MF amplitudes of the default clusters. This is shown in Figure 2.25 and shows far more detail. In simulations, where truth information is available, the SZ signal can be quantified in different ways. In Figure 2.25 the dashed lines show the normalized SZ signal Y_{500} cylindrically integrated within R_{500} while the solid lines show the normalized central SZ profile signal y_0. They clearly behave differently. Y_{500} is relatively constant as a function of mass. No mass dependence is expected from the profiles themselves and the small amount of variation can be attributed to projection effects which are more prominent at low mass. This normalized signal is larger for morphologically disturbed clusters (green dashed line) than for cool core clusters (dashed blue line) or default clusters (a value of unity) which means that the disturbed cluster signal, which is weaker in the cluster center, more than compensates for it at larger radii. The central signal, y_0 behaves quite differently. It exhibits a large degree of mass and redshift dependence due to the map resolution. High mass and low redshift clusters are better resolved and thus the central pixel averages over a smaller angular scale and suppresses the peak of the profile less. In this case the cool core clusters (solid blue line) have a stronger y_0 signal than default (a value of unity) or morphologically disturbed (solid green line) clusters, the opposite of Y_{500} since cool core clusters have more concentrated profiles.

Both y_0 and Y_{500} shown in Figure 2.25 are based on the truth. The observable MF signal, shown as blue circles for cool core clusters and green squares for disturbed clusters, can be interpreted as being in between the two extremes. The MF involves the convolution of an expected SZ pressure profile with the measured maps. This convolution effectively integrates over an aperture similarly to Y_{500}. However, the

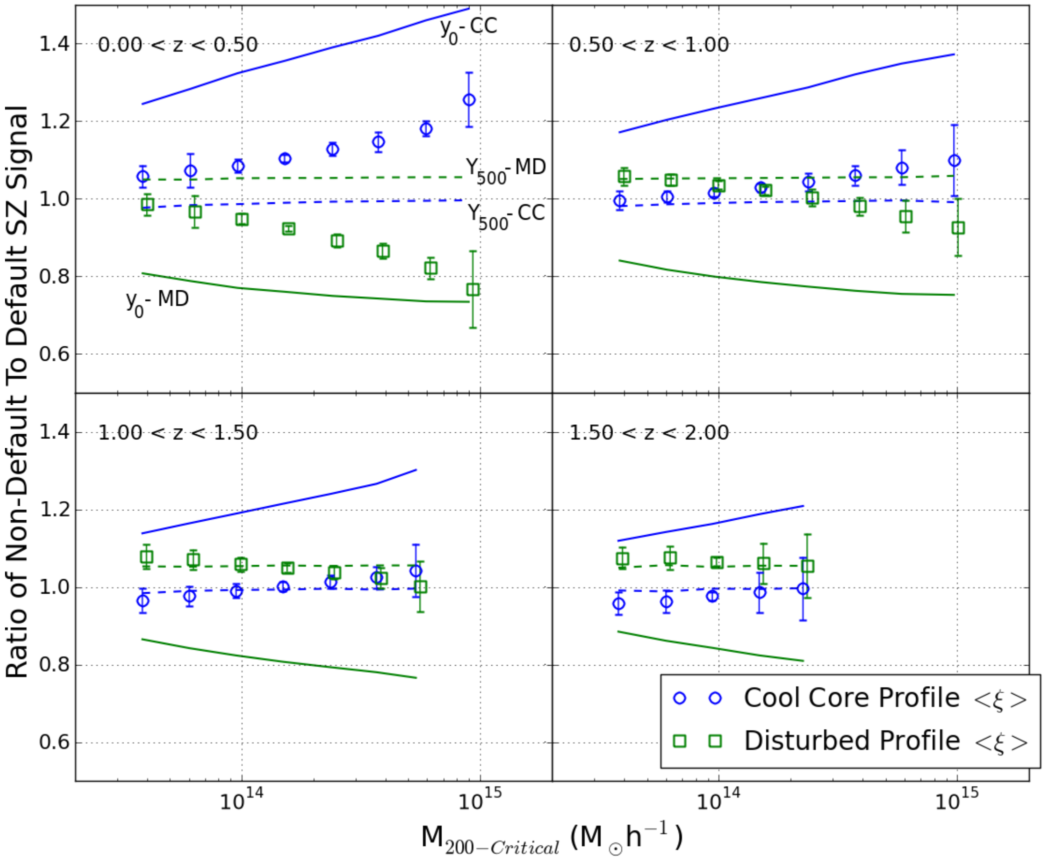


Figure 2.25: Ratios of the mean MF amplitudes relative to the default profile shape amplitudes vs mass of stacked halos in 2500 sq. degree *SPT*–like simulations. The blue circles are the ratios of cool core clusters MF amplitudes to default cluster MF amplitudes (ratios of blue circles to black × symbols in Figure 2.24). The green squares are the ratios of morphologically disturbed clusters MF amplitudes to default cluster MF amplitudes (ratios of green squares to black × symbols in Figure 2.24). The solid lines are the ratios of the profile center pixels values y_0 relative to the default profile y_0 and the dashed lines are the ratios of the integrated cluster Y_{500} to the default profile Y_{500} (blue for cool core and green for disturbed). Y_{500} and y_0 are labeled in the top left panel for clarity with "CC" standing for cool core and "MD" standing for morphologically disturbed.

MF also weights the center of the profile higher than the outskirts which is more reminiscent of measuring y_0. This explains the data points in Figure 2.25. First, we must briefly discuss the error bars shown. They were computed using bootstrap resampling; halos were removed randomly (with replacement) in each mass and redshift bin, their mean was computed and the width of the resulting distribution was used. Hence, the error bars give an accurate portrayal of the uncertainties in a given stack over the area of the simulated sky. In addition, in a real survey, one would probably use narrower redshift bins which would increase the size of these error bars. However, the halo properties, the noise model and realization were identical for the three classes of profile shapes so the *relative* differences between the means when using different gas pressure profiles are accurate. Put differently, when a cool core clusters (blue circle) is above a disturbed one (green square), its mean is actually larger even if the difference between the two points is smaller than their error bars.

We can now interpret the results. At low mass and higher redshift, the disturbed clusters show a larger signal. In these cases the pressure profiles are not resolved and the signal is more similar to an integrated Y_{500} where the disturbed halos (green dashed line) have a higher signal. At higher masses and lower redshifts, the cool core amplitude becomes larger than the disturbed cluster amplitude. This occurs for the entire mass range in the lowest redshfit bin and happens for masses larger than $1.5 \times 10^{14} h^{-1} M_\odot$ for redshifts between 0.5 to 1.0 and larger than $4 \times 10^{14} h^{-1} M_\odot$ for redshifts between 1.0 to 1.5. At redshifts higher than 1.5 no clusters are resolved sufficiently for the cool core signal to be higher than the disturbed cluster signal. Here the MF amplitude ratios line up with the Y_{500} dashed lines which indicates that the MF is integrating the profiles without any sensitivity to the internal structure.

This section illustrates the utility of HaRMLSS. It enables theoretical studies using input SZ maps (solid and dashed lines in Figure 2.25) as well as full observable studies. The conclusions we can reach here are that the three different profile classes we used can lead to significant differences in the observed signal at a fixed mass as well as changing the MF – mass scaling slope from the "true" SZ – mass scaling. At the same time though, by simulating the appropriate survey area and bin sizes, it allows us to determine the conditions when the statistical uncertainty dominates systematic as well as how different experiments will differ. For instance, *Planck* with its much larger beam size tends to integrate more of the profile and reduce differences between cool core and disturbed clusters.

2.5.2 Scaling Relation vs. Mis-centering Errors

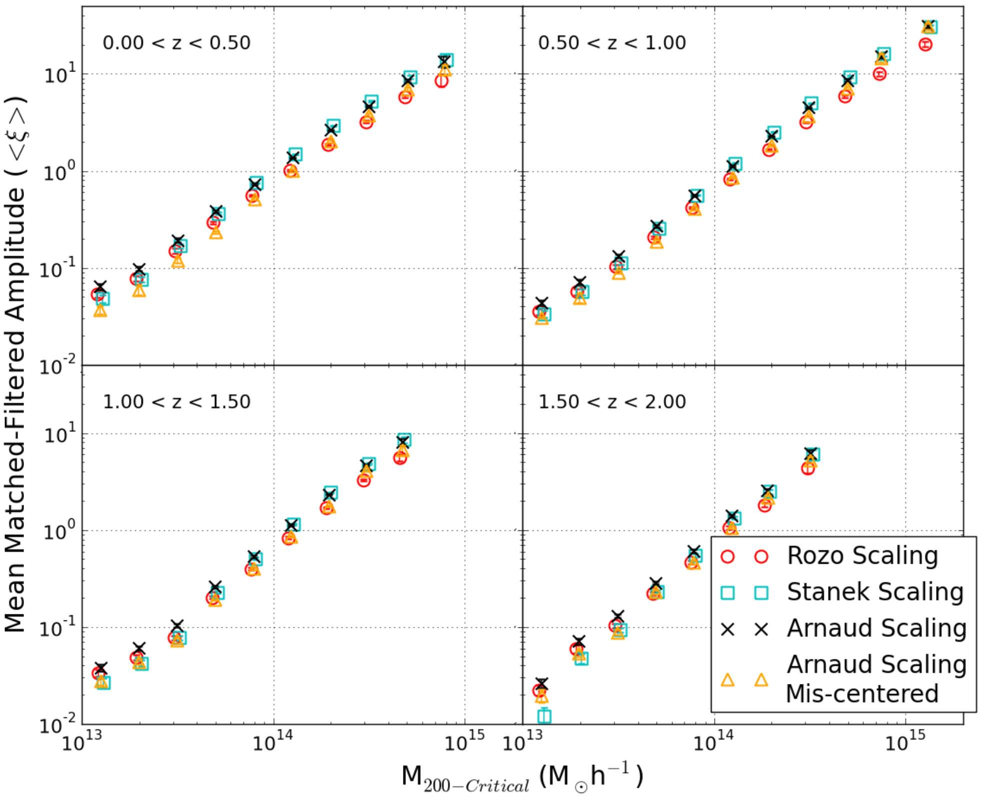


Figure 2.26: Mean MF amplitudes vs mass of stacked halos in 2500 sq. degree *SPT*–like simulations where the SZ-mass scaling was varied. The four panels represent the four redshift bins that our sample was divided into. Rozo et al. (2012) clusters are represented by red circles, Stanek et al. (2010) clusters are represented by cyan squares and the default Arnaud et al. (2010) clusters are represented by black × symbols. In addition, the Johnston et al. (2007) mis-centering model was applied to the Arnaud et al. (2010) scaling and plotted with orange triangles.

In a similar fashion, we compare stacked MF (at θ_{Core}=0.75′) amplitude scaling with mass for three input SZ – mass scalings along with a model for mis-centering from Johnston et al. (2007). The purpose here is to demonstrate the degeneracy between these very different systematic sources as well as to discuss what can be done about it. We once again simulated three *SPT* – like maps on a sky patch of approximately 2500 sq. degrees. This time they were based on the Arnaud et al. (2010), Stanek et al. (2010) and Rozo et al. (2012) scalings. In addition, we applied an optical cluster mis-centering error (Johnston et al. 2007) on the Arnaud et al. (2010) SZ – mass scaling to compare its effects. The mean stacked signal vs mass is once

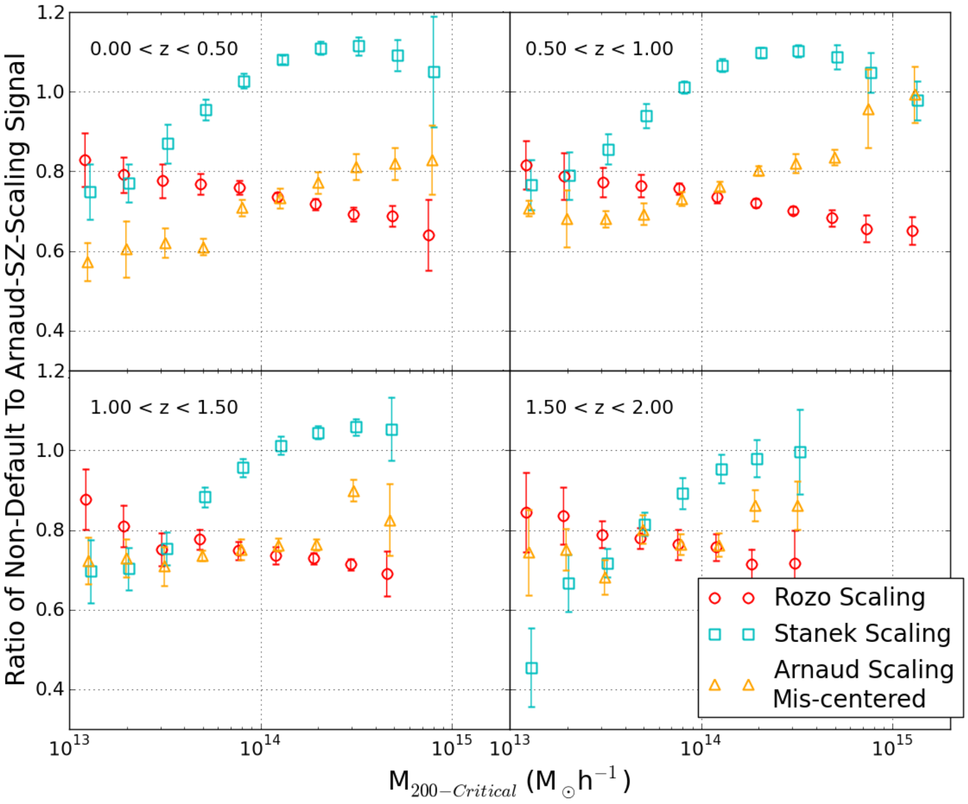


Figure 2.27: Ratios of the mean MF amplitudes relative to the default Arnaud scaling relation amplitudes vs mass of stacked halos in 2500 sq. degree *SPT*–like simulations. The red circles are the ratios of Rozo et al. (2012) cluster MF amplitudes to default Arnaud et al. (2010) cluster MF amplitudes (ratios of red circles to black × symbols in Figure 2.26). The cyan squares are the ratios of Stanek et al. (2010) cluster MF amplitudes to default Arnaud et al. (2010) cluster MF amplitudes (ratios of cyan squares to black × symbols in Figure 2.26). The mis-centering model ratio of the Arnaud et al. (2010) scalings is shown by orange triangles

again split into four redshift bins and is shown in Figure 2.26. It is not particularly informative due to the large dynamic range covered. Details of this comparison are better encapsulated in Figure 2.27. It shows the ratios of the *Stanek* (cyan squares), *Rozo* (red circles) and mis-centered *Arnaud* (orange triangles) MF amplitudes to the default *Arnaud* amplitudes. Some of the characteristics are consistent with the simulation inputs. The *Stanek* scaling exceeds the *Arnaud* scaling for masses larger $10^{14}h^{-1}\,M_\odot$ but drops below it for lower masses. The *Rozo* scaling is always below the

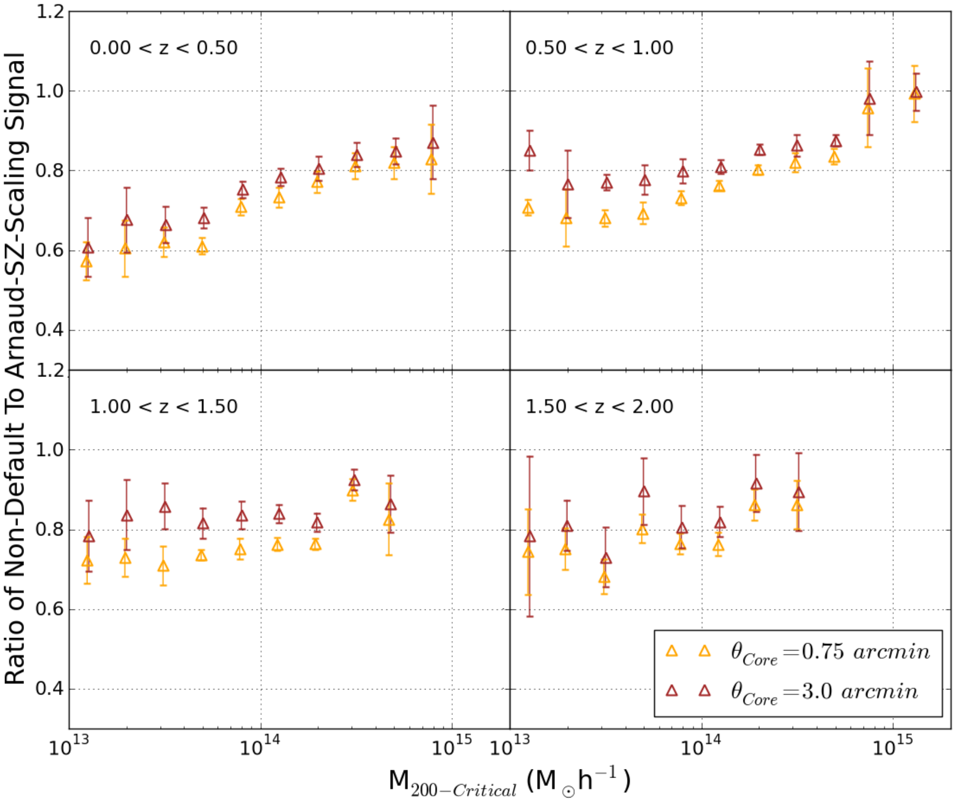


Figure 2.28: Ratios of the mean MF amplitudes of the mis-centered model using two different MF sizes relative to the default Arnaud scaling relation amplitudes vs mass of stacked halos in 2500 sq. degree *SPT*–like simulations. The mis-centering model ratio using $\theta_{Core} = 0.75'$ is shown by orange triangles and the one using $\theta_{Core} = 3.0'$ is shown using brown triangles.

Arnaud scaling though it gets more discrepant at higher masses. The mis-centering model is of interest in that it could easily be misinterpreted as a different scaling relation. However, this can be overcome by using a different signal proxy for SZ. For instance, (Biesiadzinski et al. 2012) shows that the mis-centering model has a much smaller effect. This is due to the larger size of the *Planck* beam which effectively causes the MF to integrate over a larger area and therefore be less impacted by mis-centering effects.

This is further demonstrated in Figure 2.28 where the mis-centered model from before with a $\theta_{Core} = 0.75'$ is shown as orange triangles again along with the same model using the MF with a $\theta_{Core} = 3.0'$ (brown triangles). One can see that it suffers less systematically from mis-centering errors but the gain is not very significant. The MF still weighs the center more and it suffers due to high pass filtering by the MF. Using a larger θ_{Core} or even an aperture is not sufficient to eliminate the mis-centered signal suppression. Further studies have shown that by adjusting the filtering schemes, the effects of mis-centering can be reduced even further. Unfortunately, these procedures tend to significantly increase statistical noise. This trade-off can be tested using HaRMLSS under conditions closely matching real observations in order to tune the necessary parameters as discussed in Appendix A.

2.6 Summary

We constructed a flexible and expandable SZ simulation framework called HaRMLSS for joint studies of optical and millimeter-wave data gathered to study clusters of galaxies. HaRMLSS uses the outputs of an existing N-body simulation to paint tSZ and kSZ profiles onto them. It then adds astrophysical backgrounds and foregrounds to create sky maps followed by simulating instrumental effects to end up with realistic simulated observations. The main strength of HaRMLSS is its ability to quickly realize sky maps with different SZ profiles and signal – mass scaling relations as well as the ease with which new modules can be included to make simulations more sophisticated. Many such improvements are currently planned including the use of additional N-body catalog information to better “customize” the SZ signal to dark matter halos and the use of simulated galaxy catalogs to correlate the two observables. It is also desired to implement a full gravitational lensing treatment and extend the maps to full sky with continuous primary CMB background that will enable these maps to be used for additional scientific studies.

CHAPTER 3

Impact of Optical Systematics on SZ-Optical Scaling Relation

In the previous chapter we discussed the application of the HAlo–Resolved Millimeter-wave Layered Sky Simulation (HaRMLSS) framework to look into the effects of Sunyaev–Zel'dovich (SZ) (Sunyaev & Zeldovich 1972; Birkinshaw 1999; Carlstrom et al. 2002) systematics on signal stacked at perefectly known halo locations. Realistically, this stacking will have to occur on other cluster observables such as optically detected clusters. This chapter is largely a reproduction of the previously published (Biesiadzinski et al. 2012) study of the impact of optical cluster observable systematics on stacked SZ signal in light of the *Planck* measurements.

Optical galaxy cluster surveys have identified thousands of clusters down to a mass limit of $\sim 10^{14}\, M_{\odot}$ (e.g. the maxBCG catalog of the Sloan Digital Sky Survey (SDSS) clusters, Koester et al. 2007a). Millimeter wave surveys have discovered hundreds of clusters using the SZ effect (e.g., Vanderlinde et al. 2010; Williamson et al. 2011; Marriage et al. 2011; Planck Collaboration et al. 2011b), albeit to a higher mass limit due to instrumental noise. Using these catalogs, researchers apply *mass–observable relations* to relate the true underlying halo mass to observed properties, like the galaxy member count in the optical (richness, N_{gal} or N_{200}) or the SZ decrement (y or Y_{500}).

One can probe scaling relations down to masses below the detection limit by stacking the signal around known clusters. For instance, stacking has been used to the great benefit of mass calibration in weak lensing and X-ray studies (Sheldon et al. 2009; Rykoff et al. 2008). Similarly, the SZ/X-ray cluster scaling-laws and pressure-profiles were evaluated by Komatsu et al. (2011) and Melin et al. (2011), who stacked the SZ signal from *Wilkinson Microwave Anisotropy Probe* (*WMAP*) data around known optical/X-ray clusters. These joint optical/X-ray/SZ analyses allow researchers to take advantage of the large volumes and mass ranges from optical cluster catalogs in combination with the lower scatter in the mass observable relation in X-ray/SZ catalogs (Shaw et al. 2008; Nagai 2006; Rasia et al. 2011; Motl et al. 2005).

The SZ signal recovered from stacking *Planck* data at positions of the maxBCG (Koester et al. 2007a) clusters shows a deficit of SZ signal compared to what is expected from current mass-richness scaling relationships (Planck Collaboration et al. 2011d); this discrepancy has been confirmed using *WMAP* data (Draper et al. 2012). This discrepancy manifests itself differently for two mass-richness calibrations (Johnston et al. 2007; Rozo et al. 2009a) both of which are based on the Sheldon et al. (2009) stacked weak-lensing mass measurements of the maxBCG clusters. For the Johnston et al. (2007) calibration, a simple reduction in the global weak-lensing mass calibration by 25% would eliminate the discrepancy. The Rozo et al. (2009a) mass calibration requires a larger correction and a scaling law that is not self-similar. Planck Collaboration et al. (2011d) also show that a subset of the maxBCG clusters with measured X-ray luminosities from the Meta-Catalog of X-ray Detected Clusters of Galaxies (MCXC) catalog (Piffaretti et al. 2011) can match the predicted Y_{500} vs. richness scaling relationship, although they did not consider selection effects inherent in such a hybrid catalog.

The Planck Collaboration et al. (2011d) analysis was based on a comparison of the observed Y_{500} around maxBCG clusters to two models with different mass-richness calibrations and without including optical systematics. They evaluated the impact of impurities in the optical catalog as well as scatter in the mass-richness relations and concluded that neither could account for the observed discrepancy individually. Here, we broaden the Planck Collaboration et al. (2011d) analysis to include the *uncertainties in the mass calibrations* as well as the *combined systematic effects* in optical cluster catalogs. Instead of two model predictions to compare against the data, we look at the family of predictions which come from uncertainties in the calibrations and the ranges of systematics in optical cluster catalogs.

There are numerous systematic effects in optical galaxy cluster catalogs. These include the cluster selection (as a function of mass M and redshift z) which comprises: *completeness*— the probability that a true halo will be detected; and *purity*— the probability that a detection correctly identifies a halo rather than noise (e.g., Miller et al. 2005). Cluster *redshifts* estimated using photometric data are uncertain, which introduces scatter in the observed redshift. There is uncertainty in the *mass-richness* calibration as well as *scatter*. Finally, mis-identification of BCGs in the maxBCG cluster-finder produces angular offsets between true and recovered *cluster* *centers* (Johnston et al. 2007) called *mis-centering*. Centering offsets driven by other mechanisms (e.g., astrophysical: Sanderson et al. 2009) are smaller than those caused by the BCG mis-identification, and so we do not consider them in this study.

Using mock clusters taken from N-body simulations, we directly manipulate the purity, mass-scatter, scaling calibrations and their uncertainties. We then re-create the *Planck* richness stacking technique on these mock catalogs to create model Y_{500}-richness relations and compare to the *Planck* observations. In §3.1, we describe the N-body simulations and the suite of simulated optical cluster catalogs with various systematics, the mock *Planck* SZ observations, the mock X-ray observations and the stacking procedure. We then show the results of stacking the SZ signal for each systematic to explore how each systematic can individually affect the SZ signal (§3.2), and we compare to the *Planck* joint SZ-optical and X-ray analyses (§3.2.1). Throughout this paper, we assume a ΛCDM cosmology with a Ω_Λ=0.75 and H_0=0.71 unless otherwise noted.

3.1 Simulations

We begin with a simulated mass function and halo positions from an N-body lightcone. We then impose observables and realistic systematic effects to produce mock optical catalogs and then dress the halos with gas and simulate *Planck* SZ observations.

3.1.1 N-body Lightcone

To generate the mock SZ maps and galaxy catalogs, we begin with the halo positions from a large (N = 1260^3 particles, 1000 [Mpc $h^{-1}]^3$) cosmological dark matter simulation. Cosmological parameters were chosen to be consistent with those measured from the five-year *WMAP* data (Dunkley et al. 2009) combined with large-scale structure observations, namely $\sigma_8 = 0.8$, $\Omega_M = 0.264$ and $\Omega_b = 0.044$. The simulation was carried out using the tree-particle-mesh code of Bode & Ostriker (2003). In total, the lightcone covers a single octant on the sky (~ 5000 deg^2) to a redshift of 3, containing halos with masses $M_{\rm FOF} > 3 \times 10^{13} h^{-1} M_\odot$.

The simulation does not provide any observables (e.g., richness or SZ/X-ray luminosity). We do not use the halo masses output from this particle simulation directly but rather use the procedure described in §3.1.2. With the mass resolution available from this simulation we can reproduce the properties of the maxBCG catalog, including systematics, for clusters with $M_{500} > 6 \times 10^{13}\ h^{-1}\ M_\odot$ or $N_{200} > 20$.

3.1.2 Simulated Halo Catalogs

In the *Planck* analysis (Planck Collaboration et al. 2011d), the clusters are binned according to their optical richness (N_{200}, N_{Gals}). Richness is defined as the number of bright red galaxies (within the E/S0 ridgeline) inside R_{200} that are brighter than 0.4 L^* (Koester et al. 2007a). Recall, that richness is an observed quantity and at any fixed value, clusters can have a range of true masses (the *mass scatter*).

The halo catalog provides a mass function and large-scale structure according to our chosen cosmology and similar to the observed universe. We cannot directly assign richnesses to these halo masses that match the observed scatter. Therefore we create a mock catalog of masses and richnesses and assign them to the N-body halos to preserve the large scale structure of the universe.

For each richness we center a Gaussian probability distribution function (PDF) in $ln(Mass)$. The center of the Gaussian is taken from a particular scaling relation; the width represents the scatter in $ln(Mass)$ at fixed richness. We draw from these Gaussian PDFs to create a list of masses including scatter for each richness. We adjust the number of draws from each PDF to reproduce the halo mass function (e.g., we draw more times from the low mass bins). This provides a table of richnesses and associated masses with the same halo mass function as the N-body simulation. We sort the N-body halos and this table by mass. We associate the positions of the N-body halos to the drawn table based on this ordering. This produces our mock cluster catalog which includes large-scale structure and reproduces a particular choice of scaling relations and scatter. We use these masses to create the SZ profiles (§3.1.4) and X-ray luminosities (§3.1.5).

3.1.3 Optical Cluster Catalogs

Using the procedure described above we create mock catalogs that are modified as follows to include systematic effects:

1. **Mass-richness Calibration:** We varied the richnesses of the halos according to equation 26 (and associated uncertainties) from Johnston et al. (2007):

$$\langle M_{200} | N_{200} \rangle = M_{200|20} \left(\frac{N_{200}}{20} \right)^{\alpha_N} \tag{3.1}$$

$M_{200|20} = (8.8 \pm 0.4_{stat} \pm 1.1_{sys}) \times 10^{13} h^{-1}\, M_{\odot}$ $\alpha_N = 1.28 \pm 0.04$ or equation 4 from Rozo et al. (2009a):

$$\frac{\langle M_{500}|N_{200}\rangle}{0.71 \times 10^{14}\ h^{-1}\, M_{\odot}} = \exp(B_{M|N}) \left(\frac{N_{200}}{40}\right)^{\alpha_{M|N}} \tag{3.2}$$

$\alpha_{M|N} = 1.06 \pm 0.08_{stat} \pm 0.08_{sys}$ $B_{M|N} = 0.95 \pm 0.07_{stat} \pm 0.10_{sys}$

We look at one and two σ deviations from these mass calibrations. Masses are converted from M_{200} to M_{500} assuming an Navarro–Frenk–White (NFW) profile (Navarro et al. 1996) and mass concentrations from Duffy et al. (2008), and are relative to the critical density.

2. **Completeness:** We vary the fraction of halos in bins of redshift and mass.

3. **Purity:** We add into the halo catalogs an additional number of false halos in bins of mass and redshift to create samples with different purities. We either vary the purity as a constant with mass and redshift or match the published maxBCG purity of Koester et al. (2007a).

4. **Redshifts:** We scatter the true halo redshifts by normal distributions with varying widths as large as $\sigma_z = 0.05$.

5. **Center Offsets:** We offset the center for a fraction of the clusters according to Equation 10 in Johnston et al. (2007). For the offset clusters, the actual amount of the offset is described by Eq 8 in Johnston et al. (2007). See also Figures 4 and 5 in Johnston et al. (2007).

6. **Mass Scatter** We vary the width of the log-normal distribution of masses at fixed richness.

Realizations of mock optical cluster catalogs are created to investigate the impact of individual systematic effects. These include maxBCG-like systematics (Koester et al. 2007a,b; Johnston et al. 2007; Rozo et al. 2009a) and more general systematics that are constant in redshift and mass (or richness). We also create catalogs combining maxBCG systematics to compare to data. In our maxBCG-like mocks, the fraction of incorrectly centered clusters ranges from 12% in the highest richness bins to 39% in the lowest richness bins with a mean offset of 0.6Mpc corresponding to $3'$ for a cluster at the mean redshift of z = 0.2; the completeness and purity are >90% above $M_{500} > 1 \times 10^{14}\ h^{-1}\, M_{\odot}$ and have an estimated uncertainty of 2.5%; the mass scatter is 0.45 ± 0.10, similar to *Rozo* ($\sigma_{\ln(M)|N_{200}} = 0.45^{+0.20}_{-0.18}$ (95% CL) at $N_{200} \approx 40$).

3.1.4 Mock SZ Sky Maps and the Stacked Signal

The simplified millimeter-wave simulations were generated using an earlier version of HaRMLSS (Chapter 2). Briefly, the halo SZ signals are generated using a thermal pressure profile suggested by (Arnaud et al. 2010) and used in the *Planck* maxBCG stacking analysis (Planck Collaboration et al. 2011d). We project the profile along the line-of-sight to produce a Compton-Y profile, scaled to the appropriate size for each halo redshift. Mock *Planck* observations were created in each frequency band using the appropriate beam sizes, instrument noise and primary Cosmic Microwave Background (CMB) (Planck Collaboration et al. 2011b) temperature anisotropy. We concluded that the 143 GHz channel reproduced the dominant features of the multi-frequency analysis, and so we restricted our analysis solely to this channel, which has a beam size of $7.18'$ FWHM and a noise of $0.9\ \mu K-$degree.

At the position of each optical cluster, we extracted the integrated thermal SZ signal Y_{500} from each SZ sky map using a Matched Filter (MF) (e.g., Herranz et al. 2002; Melin et al. 2006) with an Arnaud profile (Arnaud et al. 2010), the size of which is inferred from either the *Johnston* or *Rozo* richness-mass scaling relations (the same as used in Planck Collaboration et al. 2011d). We stacked these match filtered signals in richness bins; then the amplitude is calibrated by comparing the spherical Y_{500} of the halos with the amplitude in the stacked SZ signal in the absence of systematics. We found that including an intrinsic random scatter of 25% in $Y_{500} - M_{500}$ (Shaw et al. 2008) did not affect our results beyond increasing statistical uncertainties in individual catalog realizations and so we did not include this additional scatter in the following analysis.

3.1.5 maxBCG – MCXC Subsample

The *Planck* team studied a subset of the maxBCG catalog whose positions were matched to within $\sim 3'$ and 0.05 redshift separation from X-ray clusters from the MCXC catalog (Piffaretti et al. 2011). These objects can be roughly subdivided into X-ray bright, ROSAT All-Sky Survey (RASS)–based (Voges et al. 1999) clusters (blue circles in left panel of Figure 3.1) composed primarily of Northern ROSAT All-Sky (NORAS) (blue stars) (Böhringer et al. 2000) catalog objects, and into dimmer, non–RASS based (green circles) clusters with most objects from the 400 Square Degree ROSAT PSPC galaxy cluster survey (400SD) (Burenin et al. 2007) catalog (green stars).

To generate our X-ray sample we start by assigning X-ray luminosities (L_X) and

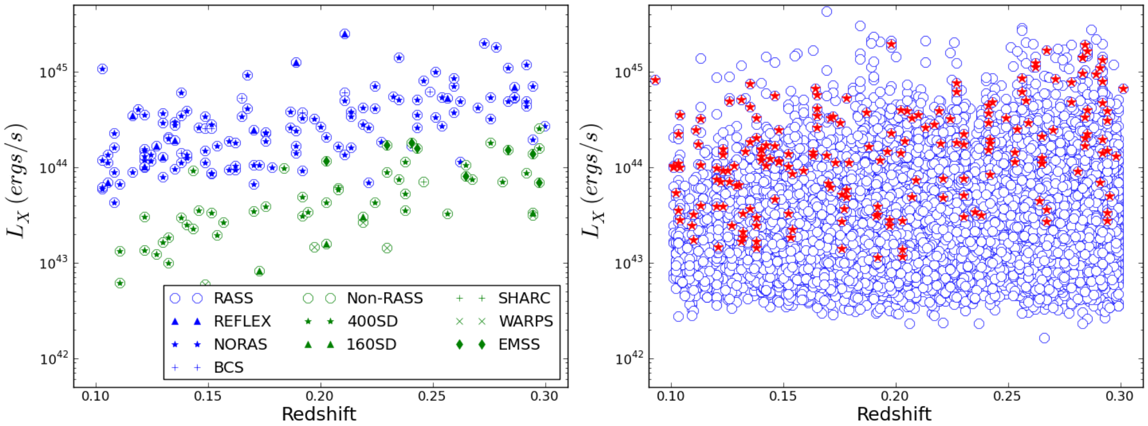


Figure 3.1: The joint maxBCG-MCXC sample selection. Left panel: The MCXC clusters correlated with maxBCG clusters. The blue markers represent the RASS–based X-ray clusters and the green symbols represent non–RASS based clusters. Right panel: Our simulated sample X-ray selection (red stars) from among all halos (blue circles).

scatter according to Table 1 in Arnaud et al. (2010) to our simulated halos based on their masses. We reproduce the scatters in the L_X scaling relations (at fixed mass and richness) that are observed in Rozo et al. (2009a) where $\sigma_{\ln(L)|M}$ ranges from 0.5 at low mass to 0.45 at high mass and $\sigma_{\ln(L)|N}$ is a constant 0.85 at all richnesses. We also vary the input L_X and scatter to assess the sensitivity to those parameters. One such simulated realization is shown as blue circles in the right panel of Figure 3.1. We then select subsets of the simulated halos, shown as red stars in the right panel of that figure, which have the same redshift and L_X distribution as the MCXC subsample. This allows us to reproduce the MCXC subsample without needing to characterize the exact selection function, which is undoubtedly complex as this catalog is drawn from heterogeneous X-ray data. We also ensure that the mis-centering for this mock MCXC-maxBCG catalog is truncated at $3'$. The maxBCG-MCXC mock catalogs need not have the same scatter in the mass-richness relation as we imprinted into the full maxBCG mock samples. This is because we imprint the observed scatter from Arnaud et al. (2010) directly onto the full catalog and then draw a sub-sample. For the MCXC/maxBCG mock subsamples, $\sigma_{\ln(L)|N}$ drops to 0.70 and $\sigma_{\ln(M)|N_{200}}$ drops to 0.40.

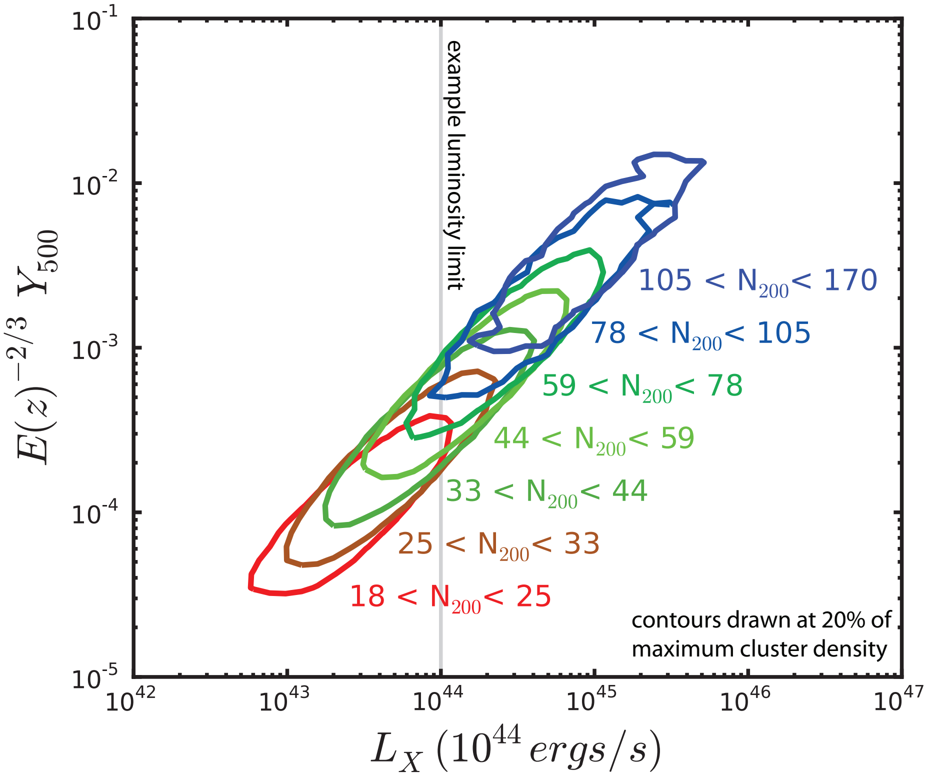


Figure 3.2: Correlation between simulated X-ray luminosities (L_X) and SZ signals (Y_{500}) in the richness bins used in this work. The contours are drawn where the number of clusters is 20% as large as the number at the mean value of L_X and Y_{500} for each richness bin (the center of each contour). The gray vertical line illustrates the approximate limit in X-ray luminosity reached by some of the surveys used in the construction of the MCXC (Piffaretti et al. 2011) where they overlap with maxBCG.

3.1.6 Correlations of Observables

The large scatter in true mass at fixed richness (see §3.1.3) induces a correlation between the observed X-ray luminosity and SZ signal. We note that this correlation is distinct from a secondary correlation in the scatter of observables. Figure 3.2 shows the cluster density as a function of L_X and Y_{500} in various richness bins. Within each richness bin there is a strong correlation between the two observables. This will be crucial for understanding the joint maxBCG-MCXC subsample discussed in §3.2.2.

Our simulation pipeline does not create correlated scatters in the observables at fixed mass. Such correlations are expected due to common substructue within clusters and projection effects (White et al. 2010). They are likely secondary effects beyond the scope of this work (see Angulo et al. (2012)).

3.2 Results

In Figure 3.3, we compare the stacked Y_{500} in our family of mock cluster catalogs to a "perfect" cluster catalog that has been calibrated according to Rozo et al. (2009a). The "perfect" catalog uses a single calibration and does not contain any of the systematics we discuss in §3.1.3. This is identical to the model the *Planck* team used to compare to the data (Planck Collaboration et al. 2011d). In each panel, the solid black line shows the average ratio (over multiple mock realizations) for models which apply the fiducial maxBCG values for calibration, mis-centering, purity, and mass scatter individually (as described at the end of §3.1.3). The gray bands show the range of models using the 1 and 2 σ uncertainties on those parameters. Dotted-lines show more general models (e.g., 70% purity independent of mass). We also show the *Planck* data presented in Planck Collaboration et al. (2011d).

Systematic uncertainties ($2\,\sigma$) in the mass-richness calibration result in up to 50% range in the model Y_{500} measurements. This is because the Y_{500} values from our perfect catalog are calculated from a single mass calibration, while the model Y_{500}s are calculated using the masses drawn from the calibration including 1 and 2 σ uncertainties.

Mis-centering suppresses (biases low) the model Y_{500} over the entire mass range, with the largest effect at low mass ($\sim$ 25% suppression). This can be understood from the convolution of the *Planck* beam ($\sim 7'$ full width at half of maximum) and the centering offsets which are on average $\sim 3'$ at the median redshift of the optical sample. The offsets are large compared to the *Planck* beam, which blurs out the SZ-signal after the convolution. The impact of this effect increases to $\sim$ 25% at low mass, since the maxBCG mis-centering fraction is mass dependent.

Impurities suppress (biases low) the amplitude of the model Y_{500}s by introducing pure noise into the SZ maps. As also noted by Planck Collaboration et al. (2011d), high levels of impurity would be required to explain the discrepancy with the data. Just as important, the weak-lensing calibration of the mass-richness relation would also be affected by large impurities which would lead to an enhancement in the mass-richness relation. Since $Y_{500} \sim M^{\frac{5}{3}}$, high impurities could even cause the observed SZ signal to be enhanced compared to the systematics-free case (something neither we nor *Planck* detect). Accurate modeling of the impact of impurities on Y_{500} requires simulating its effect on the weak-lensing calibration of the optical catalog.

The stated uncertainty in mass scatter (Rozo et al. 2009a) does not have a significant impact on the SZ signal recovered using a maxBCG-like catalog (see the gray

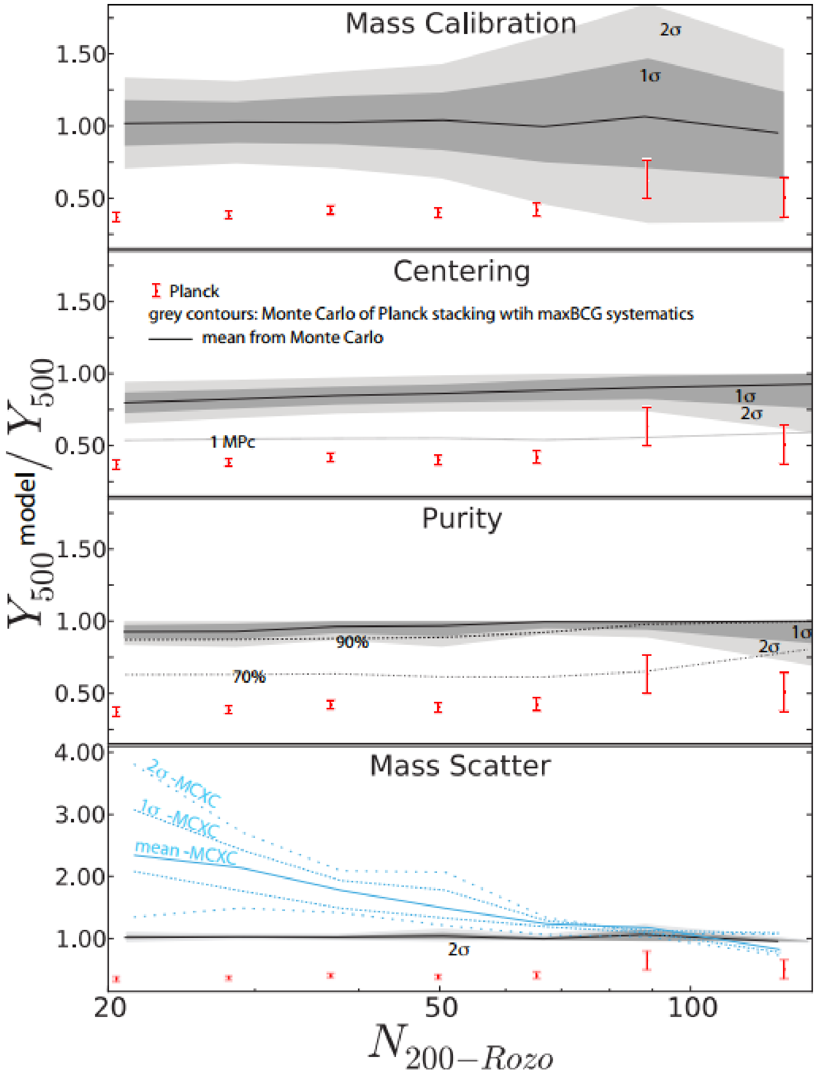


Figure 3.3: A comparison of the stacked Y_{500} in our family of mock cluster catalogs to a single "perfect" cluster catalog that has been calibrated according to Rozo et al. (2009a). The solid black lines show the model with maxBCG-like systematics included (individually). The gray bands show the range of models after we include the 1 and 2 σ uncertainties on the individual optical systematics in addition to statistical uncertainties. Gray dotted lines show more general models, while the blue lines in the bottom panel are specific to the maxBCG/MCXC sub-sample. The red error bars are the *Planck* data. Uncertainty in the mass calibration is the dominant effect on the model predictions, however impurity and mis-centering both bias the model predictions towards lower values of Y_{500}. On the other hand, X-ray luminosity selected sub-samples (e.g., the MCXC) show highly biased Y_{500} predicted values (compared to a perfect optical catalog). See Figures 3.5 and 3.6 for the combined effects of these systematics.

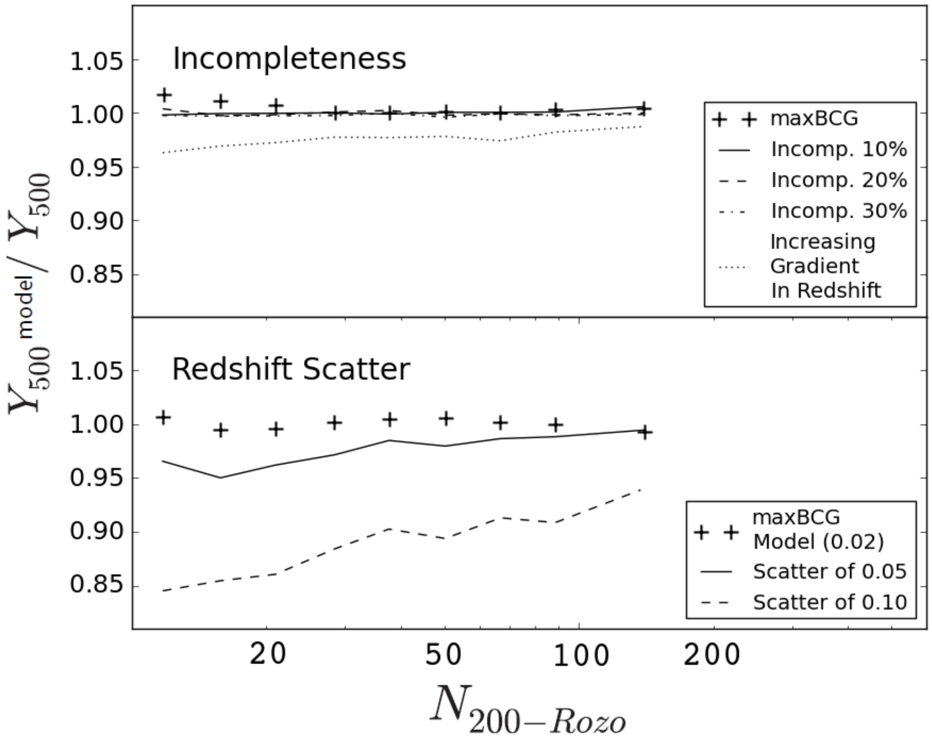


Figure 3.4: An additional comparison of the stacked Y_{500} in our family of mock cluster catalogs to a single "perfect" cluster catalog that has been calibrated according to Rozo et al. (2009a). The black plus signs show the model with maxBCG-like systematics included (individually). Top panel: effects of incompleteness. Black solid, dashed and dashed-dotted lines represent incompleteness of 10%, 20% and 30% respectively. Dotted black line represents a model where completeness improves with increasing redshift. Bottom panel: effects of redshift scatter. The solid and dashed black lines show the effect of 0.05 and 0.10 scatter in redshift, respectively.

band in the bottom panel of Figure 3.3). However, the same can not be said for the MCXC-like subsample (blue lines in the same panel and see §3.1.5). The X-ray selection causes a Malmquist bias in low richness bins where the X-ray sub-sample preferentially contains brighter (and thus more massive) clusters. Figure 3.2 illustrates that selecting clusters above some L_X limit (like the example shown by the gray line) preferentially selects clusters with high Y_{500}. Larger mass scatter increases the correlation between L_X and Y_{500} and therefore enhances the Malmquist bias. Richness bins that lie completely to the right of the L_X limit are not affected by this bias and so the SZ signal there is not enhanced.

We investigated additional systematic effects, redshift scatter and catalog completeness, and found them to have little effect with the stated maxBCG parametrization and uncertainties. These results are shown in Figure 3.4. We found that com-

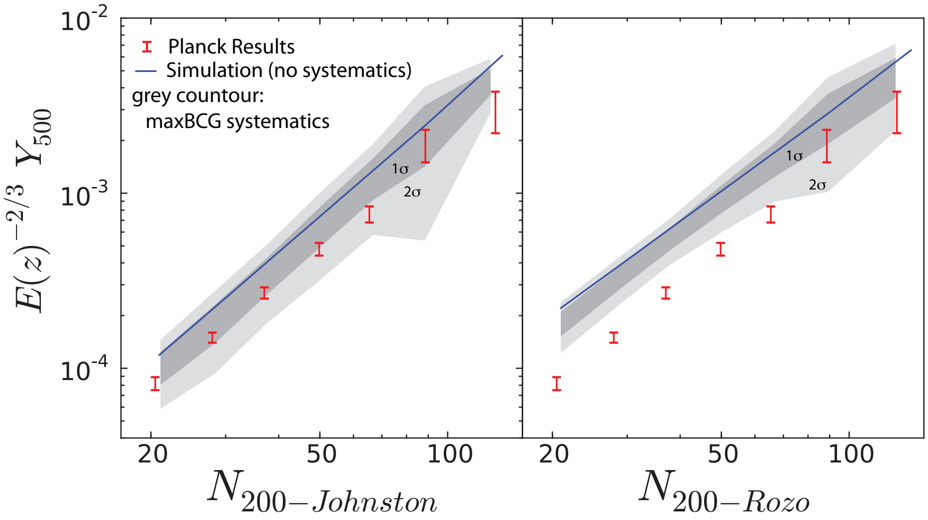


Figure 3.5: The *Planck* data (error bars) compared to the single perfect model used in (Planck Collaboration et al. 2011d) (blue line) and to the range of models (gray bands) after jointly combining all of the individual systematic effects seen in Figure 3.3. The naive perfect model predicts higher (on average) Y_{500} values compared to the models which include catalog systematics. The data are consistent with our model predictions within $1\,\sigma$ for the *Johnston* mass calibration.

pleteness is only important if it changes as a function of redshift. This is demonstrated in the top panel of Figure 3.4 where setting incompleteness to 10%, 20% or 30% has no effect. However, a gradient where completeness increases in redshift (dotted line in the figure) suppresses the signal. This is due to the self-similar scaling of Y_{500}. The maxBCG-like simulation (plus signs in the figure) show a slight excess of SZ signal at low richness, because completeness gets a little worse there at higher redshifts. In addition, redshift scatter will not suppress recovered SZ signal until it becomes larger than 0.05 (bottom panel of Figure 3.4).

3.2.1 Simulating *Planck* – maxBCG Joint Analysis

Figure 3.5 compares the *Planck* results to our models. The *Planck* data (error bars) are the same in both panels from Planck Collaboration et al. (2011d). The solid blue lines shows the single naive perfect model based on either the *Johnston* (left) or *Rozo* (right) mass calibration in the absence of systematics (Planck Collaboration et al. 2011d). The gray bands show model predictions based on our Monte-Carlo mock cluster catalog realizations which include all of the maxBCG optical catalog properties, uncertainties, and systematics shown in Figure 3.3 and which were applied

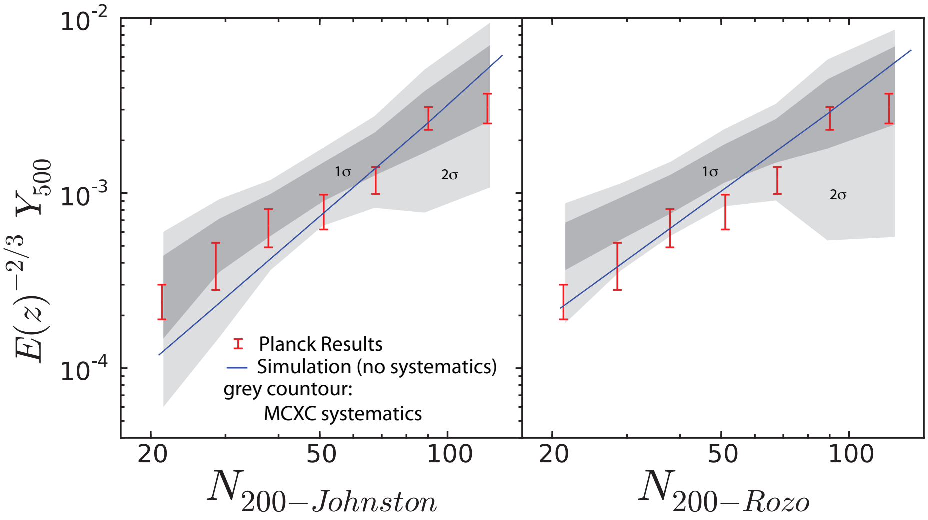


Figure 3.6: The *Planck* data for the maxBCG/MCXC X-ray sub-sample (error bars) compared to the single perfect model used in (Planck Collaboration et al. 2011d) (blue line) and to the range of models (gray bands) after jointly combining all of the individual systematic effects seen in Figure 3.3. While the perfect model is the same as in Figure 3.5, the gray bands here include the bias seen in Figure 3.3 (bottom), which is caused after sub-sampling clusters based on their X-ray luminosities to match the observed data. The naive perfect model predicts lower (on average) Y_{500} values compared to the models which include catalog systematics. The data are consistent with our model predictions at the 1 (2) σ levels on the optical systematics for the *Johnston* (*Rozo*) mass calibration.

in the original weak-lensing richness mass calibrations. While the *Planck* data are statistically inconsistent with the naive perfect model prediction, they lie at the lower edge of the models which include the $\sim 1\,\sigma$ systematic uncertainties for the *Johnston* mass calibration.

3.2.2 Simulating the maxBCG-MCXC Joint Sample

Figure 3.6 shows our prediction for the MCXC sub-sample of the maxBCG catalog compared to the *Planck* data. The gray bands here include simulated optical and X-Ray systematics as well as the X-ray selection function. As expected from Figure 3.3-bottom, we see a bias in the predicted Y_{500} with decreasing richness due to the Malmquist bias present in low richness bins after only the brightest L_X are selected (see Figure 3.2). The *Planck* observations lie inside the lower edge of the models which include the 1 and 2 σ systematic uncertainties for *Johnston* and *Rozo* mass calibrations respectively.

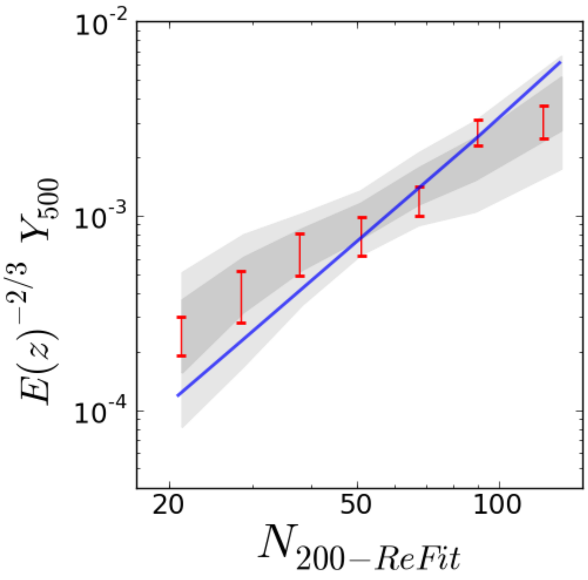


Figure 3.7: The *Planck* data for the maxBCG/MCXC X-ray sub-sample (error bars) compared to the single perfect model used in (Planck Collaboration et al. 2011d) (blue line) and to the range of models (gray bands) that accounts for the uncertainty in the re-fit scaling relation and X-ray selection effects and systematics.

Recall previously (Figure 3.5) that there is evidence for a systematic underestimate of the weak-lensing mass-richness calibration. We thus re-calibrate the mass-richness relation using the Y_{500}-richness data in Figure 3.5. By definition, this brings the Y_{500}-richness relations into full agreement for the full maxBCG samples. We then apply this new calibration to predict the Y_{500} values for the MCXC-like mock sub-samples and show the result in Figure 3.7. The re-calibrated prediction is in excellent agreement with the data, somewhat better than *Johnston* mock-MCXC sample calibration. In other words, we can use the full maxBCG-SZ data to “fix” the optical calibration and simultaneously achieve better agreement between the predicted Y_{500}s and real data. This of course completely ignores the contribution of the SZ systematic effects and should not be taken at face value. However, this approach does show the strength in using multiple observables while studying clusters of galaxies.

3.3 Discussion

The *Planck* team reported that the stacked SZ signal around optical clusters lies well below the single model expectation which does not include the optical catalog systematic uncertainties. On the other hand, they find that the observed stacked Y_{500} values around an X-ray limited sub-sample are consistent with the naive optical model. They concluded that the gas properties of clusters appear to be more stably related to each other than the gas-to-optical properties of clusters (Planck Collaboration et al. 2011d). In this work, we reach a fundamentally different conclusion: the

Y_{500} values observed by *Planck* are consistent with the model predictions for both the entire cluster sample and the X-ray sub-sample to within the $1\,\sigma$ optical systematic uncertainties of the Johnston et al. (2007) mass calibration. Not only do we argue that there is no significant discrepancy between the models and the observed *Planck* stacked Y_{500} values around optical clusters, but we also argue that the optical and X-ray selected sub-samples simultaneously agree with model predictions. For instance, we can apply a single mass-richness calibration to the data and fit the predicted Y_{500} models in Figures 3.5 and 3.6 simultaneously. We find that the dominant source of optical systematic uncertainty comes from the mass calibration, which alone can account for most of the original discrepancy noted by Planck Collaboration et al. (2011d). Impurities and centering errors combine to bias the model predictions towards lower Y_{500} for the optical samples while mass scatter biases the predictions high for low richness systems in the X-ray limited subsample. When fully accounted for, these systematics allow for models which are matched by the observed data for both the optical and X-ray cluster sub-samples in the *Planck* data. The range on the acceptable models is quite large and we note that the SZ-optical scaling laws cannot by precisely characterized using this type of stacking until the optical systematics improve (specifically mass calibration and its scatter).

This work highlights the importance of multi-wavelength studies of cluster properties as a source of cross-checks and a calibration. It is clear that optical systematics cannot be ignored and future analysis of stacked clusters should be done using Monte Carlo analysis to include a larger suite of systematic errors.

CHAPTER 4

Millimeter Wave Simulation And Application Summary

The coordinated multi-wavelength surveys of the cosmos offer a wealth of information for observers. However studying such surveys requires a great deal of preparation and planning. This is especially true for galaxy cluster cosmology since the theoretically predicted objects, dark matter halos, are not directly observable. Instead, their density, masses and evolution have to be inferred from different observables of the baryons they host and gravitational distortions they cause.

The HAlo–Resolved Millimeter-wave Layered Sky Simulation (HaRMLSS) framework discussed in chapter 2 was designed to enable Sunyaev–Zel'dovich (SZ) studies coordinate with optical and/or X-ray observations. This is accomplished by creating simulated sky patches with noise and background characteristics resembling observations performed by real observatories and signal correlated with an underlying cosmological N-body simulation of dark matter halos. This N-body simulation is also used to create mock optical observational catalogs and work is ongoing to create simulated X-ray images of the same halos. This allows for the study of a mock universe using three observables that will also be utilized in the real world. The large overlap between Dark Energy Survey (DES) optical observations, VISTA Hemisphere Survey (VHS) near-infrared (NIR) observations, the completed *South Pole Telescope* (*SPT*) and ongoing *SPT-POL* SZ observations and various X-ray observations will allow for detailed characterization of galaxy clusters and therefore for the derivation of cosmological parameters.

While HaRMLSS in its current incarnation is well suited to galaxy cluster studies there are multiple steps that must be take to make it useful for next generation CMB experiments. First and foremost, the weak lensing signature of galaxy clusters must be imprinted onto the CMB background in order to utilize HaRMLSS for CMB lensing studies that give us another means of determining the mass of clusters. In addition, the polarized SZ signals must also be introduced into the maps. These improvements will make the simulated maps more realistic. Using the full sky primary CMB background, instead of the flat sky approximation currently in use, they will become useful for a whole set of CMB experiments. Additional foreground improvements may also be implemented in the near future. The point source populations should be tied to simulated optical galaxies and realistic correlations with halos so that effects like gravitational magnification can be included. Addition of the galactic synchrotron foreground is also possible.

The existing HaRMLSS framework can already begin to address the many questions associated with SZ observations as well as our general understanding of galaxy clusters. As discussed in Chapter 3 the current mass calibration of clusters may not be as accurate or precise as is often assumed. The significance of the disagreement between SZ measurements and predictions based on optical mass proxies served to alert the community (e.g. Biesiadzinski et al. 2012; Angulo et al. 2012; Rozo et al. 2012) to the lack of a consensus in cluster characterization. This is currently being addressed by many groups who attempt to use X-ray, SZ and/or weak lensing observations simultaneously to constrain cluster mass calibrations and their uncertainties. Progress is also necessary on the hydrodynamical simulation front to better understand the common systematics that may bias mass measurements.

HaRMLSS uses relatively simple SZ profile shapes. As indicated in §2.5.1 these shapes can have a very large impact on the recovered signal. Additional shapes based on observations and simulations should be tried to determine how best to recover the SZ signal. Of immediate interest are comparisons between HaRMLSS mocks and hydrodynamical simulations. Splitting the halos generated by HaRMLSS and hydrodynamical simulations into isolated and interacting samples would allow us to better understand the fidelity of our method of SZ profile placement. It may also help us understand how gas interactions alter the SZ signal recovered via different means.

A larger scale undertaking would consist of the investigation of the various data sets shown in Figure 2.8. Various groups using different instruments converted their observables into common Y_{500} values using assumptions about profile shapes and in some cases additional X-ray or weak lensing information. It is of interest to determine if accounting for these methods of signal recovery could lead to a more uniform mass – SZ scaling relation. This is a task that HaRMLSS is especially well suited for due to its flexibility in profile shape generation and scaling relation implementation.

The large amount of data expected from DES will be of great help in understanding clusters in the near future. With large samples and sophisticated network algorithms, *SPT* observations will be stacked on optically relaxed and unrelaxed clusters to characterize the impact of the state of the gas and to determine the best (least susceptible to gas physics complications) SZ mass proxies. Similar approaches will be taken with X-ray observations. An ongoing project with the DES and *SPT* collaborations is stacking SZ observations on optically selected Luminous Red Galaxys (LRGs) with plans to extend these to galaxy clusters as more data becomes available. Various optical and SZ systematics must be understood using HaRMLSS and other tools to make such measurements meaningful. In addition, direct comparisons of high resolution SZ and X-ray images may allow observers to better quantify the impact of non-thermal physics on the gas dynamics and therefore improve our understanding of cluster masses.

CHAPTER 5

Measurements and Mitigation of Reciprocity Failure

Optical observables can be vastly improved by extending them into near-infrared (NIR). For example, the light curves of Type Ia Supernovae (SNe) (Brown 2007) and galaxy correlations observed at higher redshifts can help constrain cosmological parameters. In addition, higher redshift clusters can be detected more easily and using very high redshift galaxies would allow direct weak lensing measurements of their mass. However, NIR detectors often suffer from complicated systematic uncertainties. The dark energy instrumentation group at the University of Michigan undertook the task of measuring reciprocity failure, a particular flux-dependent non-linearity, in four Mercury-Cadmium-Telluride (HgCdTe) devices produced for the *Supernovae Acceleration Probe* (*SNAP*) effort (Aldering et al. 2002). The results summarized in this chapter have been previously published in Schubnell et al. (2010), Biesiadzinski et al. (2011b) and Biesiadzinski et al. (2011a).

NIR detector technology has made great strides over the past two decades and large format arrays with excellent performance are now commercially available. Substrate – removed devices extend the wavelength sensitivity of near infrared detectors into the UV and highly integrated read-out application-specific integrated circuits (ASICs) provide compact, low power front-end electronics. Advances in detector technology make NIR detectors well suited for space-based wide-field imaging instruments that can utilize various probes of dark energy. Most of those probes rely on photometric calibrations over a wide range of intensities using standardized stars and internal reference sources. Hence, a complete understanding of the linearity of the detectors is necessary. Reciprocity failure was observed in the Near Infra-Red Camera and Multi-Object Spectrometer (NICMOS) on the *Hubble Space Telescope* (*HST*) (Bohlin et al. 2005; de Jong et al. 2006). The NICMOS instrument, installed onboard *HST* during the second servicing mission in 1997, employs three 256×256

NIR detectors. These $2.5\,\mu$m cut-off HgCdTe devices were fabricated by Rockwell Science Center, now Teledyne Imaging Sensors (TIS). This vendor also supplied the 1024×1024 $1.7\,\mu$m cut-off HgCdTe detector for the Wide Field Camera 3 (WFC3) instrument (Baggett et al. 2008), which was recently installed on *HST* during the final servicing mission. The four $1.7\,\mu$m cut-off HgCdTe detectors used for the reciprocity study described here were also supplied by TIS.

The NICMOS team concluded that the NICMOS detectors exhibit a significant flux dependent non-linearity which strongly varies with wavelength (Bohlin et al. 2005). Additional reports of count rate dependent non-linearity observed in HgCdTe NIR detectors (Bohlin et al. 2005; Riess 2010; Hill et al. 2010; Deustua et al. 2010; Biesiadzinski et al. 2011b,a) suggest that this effect is common in HgCdTe detectors, although so far only measurements with detectors from the HgCdTe Astronomy Wide Area Infrared Imager (HAWAII) family produced by TIS, have been reported. For the NICMOS detectors a non-linearity of about $6\,\%\,\mathrm{decade}^{-1}$ was reported based on a comparison of NICMOS and Space Telescope Imaging Spectrograph (STIS) standard star observations (Bohlin et al. 2005). Measurements on $1.7\,\mu$m cutoff detectors produced for the NIR channel of the WFC3, installed onboard the *HST* during the final servicing mission in 2009, show reciprocity failure between 0.3 % and 1 % in the wavelength range from $0.85\,\mu$m to $1.0\,\mu$m (Hill et al. 2010).

Reciprocity failure must be carefully distinguished from the well-known non-linearity of total signal, referred to here as integrated-signal non-linearity, which is observed in near infrared detectors that integrate charge on the junction capacitance of the pixels. Integrated-signal non-linearity in NIR detectors is caused by dependence of diode capacitance on voltage and non-linearity in the readout multiplexer, and is usually measured by integrating a constant flux for different exposure times. Reciprocity failure in turn can be measured by varying the flux for exposure times that produce a constant integrated signal.

The mechanism responsible for reciprocity failure is not yet understood. It has been suggested that image persistence in HgCdTe detectors is caused by the slow release of trapped charge in the bulk material (Smith et al. 2008). It is conceivable that charge traps are also the cause of reciprocity failure since they would prevent charge from being collected at the pixel capacitor by absorbing it. Alternatively, this non-linearity could originate in the HAWAII multiplexer, or it may be caused by small leakage currents at the charge integrating transistors. Mathematically, reciprocity failure can be characterized by a logarithmic behavior over most of the dynamic range of a detector and the deviation from a linear system is expressed as fractional deviation per decade of total signal response.

Reciprocity failure impacts photometry as residual pixel-level uncertainties directly propagate to the estimated uncertainty on the derived magnitude. Detailed knowledge of the degree of reciprocity failure for a detector will affect the calibration strategy and the calibration devices needed. A profound understanding of the cause of this effect could influence the detector manufacturing process, possibly reducing or even eliminating this non-linearity.

5.1 Instrument

To quantify reciprocity failure in NIR detectors, a dedicated test system was designed and built. Based on the measurements reported by the NICMOS team it was determined that a sensitivity to reciprocity failure of at least 1 %/decade over the full dynamic range of a typical NIR detector had to be achieved, though a limit of 0.1 %/decade was eventually reached. To measure reciprocity failure a detector was exposed at different illumination intensities, and the incident flux was precisely monitored with photo-diodes. The exposure time at each illumination intensity was adjusted to integrate to similar total integrated signals whenever possible. A parametrization including integrated-signal non-linearity and reciprocity failure was used to describe the data and to extract a measurement of the non-linearity due to reciprocity failure (see §5.3.1). Knowledge of the linearity of the photo-diodes is essential to this method. Therefore, deviation from linearity of the photo-diodes was measured independently as described in §5.1.2.

The experimental set-up utilizes a fixed illumination geometry. The illumination intensity is varied through a combination of neutral density (ND) filters and pinhole apertures, as schematically illustrated in Fig. 5.1. A regulated light source placed outside the dewar is connected via a liquid light guide to a glass rod that illuminates a pinhole mounted on the aperture wheel inside the dewar. To avoid stray light entering the dewar, the glass rod is surrounded by a bellows that attaches to the cold shield and the aperture wheel. The detector is illuminated by an integrating sphere, placed immediately below the aperture wheel, with fixed aperture and baffling. This produces an illumination profile at the detector that is independent of illumination intensity. The baffle tube, located between the integrating sphere and the detector, prevents stray light and reflected light from reaching the detector and keeps the illuminating geometry fixed. A set of six pinhole apertures at the input of the integrating sphere combined with ND filters at the entrance of the dewar extension allow

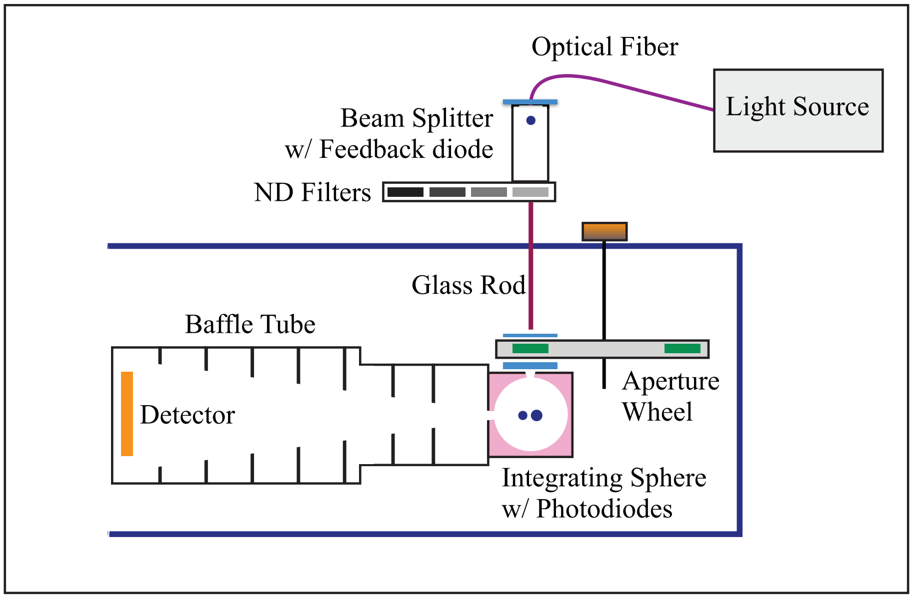


Figure 5.1: Schematic overview of the set-up used to measure reciprocity failure. Not shown is the liquid nitrogen vessel to which this set-up is attached.

a dynamic range in intensity of approximately 10^6 to be covered. Because all measurements are relative to the photo-diodes that monitor the incident flux, knowledge of the exact area of the pinholes is not critical. Furthermore, knowledge of the exact optical densities of the ND filters is also not essential. Since ND filters can show spectral dependence, pinhole apertures were used to verify the spectral flatness of the ND filters utilized in the set-up at a level sufficient for the measurements reported here.

5.1.1 Illumination

The detector inside the dewar is illuminated by one of two light sources: a feedback controlled 50 W Quartz-Tungsten-Halogen (QTH) lamp or alternatively a 790 nm diode laser. Light from the QTH light source is guided by a liquid light guide (Newport 77634) to a 70/30 beam splitter for feedback diode pick-up. A Silicon (Si) feedback diode connected to the QTH lamp control electronics stabilizes the QTH light source. Bulbs were changed frequently to avoid end-of-life fluctuations and spectral variations. A filter stack in front of the beam splitter provides for pass-band selection. Depending on the wavelength selected for the measurement, either a 900 nm long-pass filter or a stack of a 1100 nm short-pass filter and a 1000 nm short-pass filter (to improve out-of-band blocking) was inserted into the light path. The pass

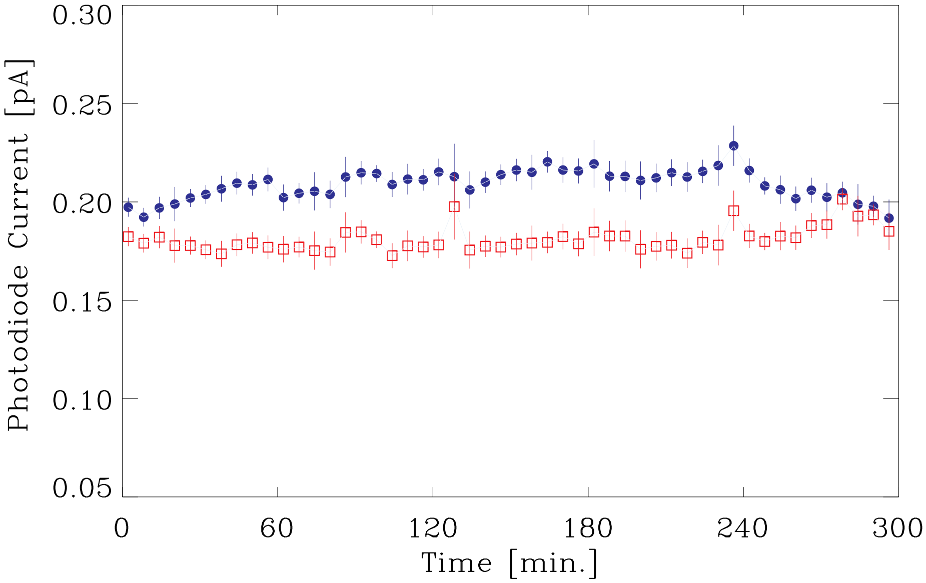


Figure 5.2: InGaAs photo-diode current as a function of time. The blue circles show the time averaged dark corrected current registered during a reciprocity measurement extending over 5 hours. The red squares show the same photo-diode measurement corrected for fluctuations of the pico-ammeter.

filter is then followed by one of four band-pass filters.[1] Following the splitter, the re-focussed light beam passes through a filter slide, housing a selectable set of ND filters with optical densities 0, 1, 2, and 3. The connection from the warm optics into the dewar is made by a glass rod. Light from the glass rod is then incident on the selected aperture inside the aperture wheel. The aperture wheel has a total of eight positions, six of which house pinholes ranging in diameter from 30 μm to 11 mm (30 μm, 100 μm, 330 μm, 1 mm, 3.3 mm, and 11 mm), one position completely blocks the light, and one position is fully open with no aperture ($\approx$ 13 mm diameter).

The pinhole illuminates the entrance port of a 2-inch integrating sphere (SphereOptics SPH-2Z-4) as shown in Fig. 5.1. An optional short-pass cold filter (Asahi YSZ1100) between two diffusers just in front of the integrating sphere is used for measurements below 1000 nm. The inside of the integrating sphere is coated with polytetrafluoroethylene based material providing good reflectivity at NIR wavelengths and good low temperature performance.

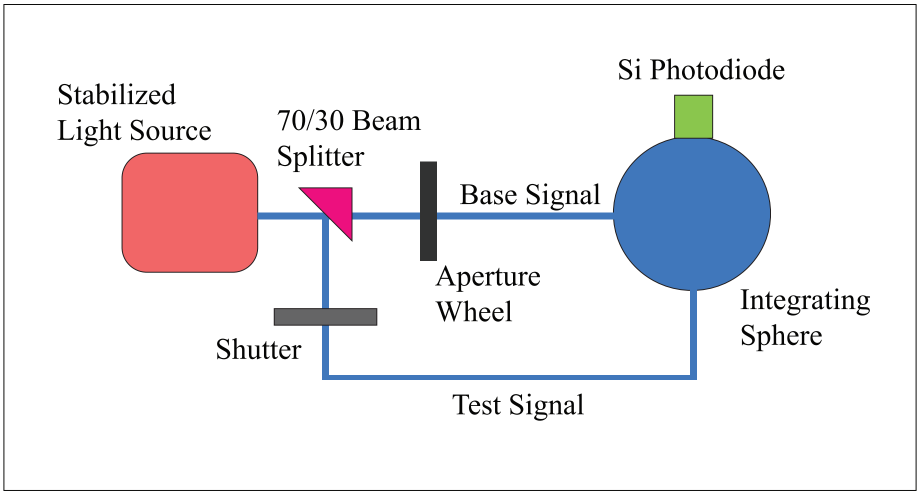


Figure 5.3: Schematic set-up used to measure Si photo-diode linearity.

5.1.2 Photo-diode Calibration

The reciprocity set-up was designed for measurement of substrate removed NIR HgCdTe detectors which exhibit spectral response at visible and NIR wavelengths. Two photo-diodes, an indium gallium arsenide (InGaAs) photo-diode and a Si photo-diode, were selected for good wavelength coverage. The NIR photo-diode is a blue extended InGaAs PIN diode (Hamamatsu Photonics G108799-01K) with an effective area of 0.785 mm^2 and spectral response range of 0.5 μm to 1.7 μm. For improved sensitivity in the visible, a Si photo-diode (Edmund Optics 53371) with an effective area of 5.1 mm^2 and spectral response between 0.5 μm and 1.1 μm was used. The two photo-diodes were mounted adjacent to each other to an open port of the integrating sphere as shown in Fig. 5.1 and were read out in parallel.

The photo-diode currents were recorded by two Keithley 6485 pico-ammeters that were read out through a GPIB interface by the data acquisition computer. For stable performance, the pico-ammeter was turned on at least 1 hour prior to every series of measurements. Typical photo-diode currents were of order 1 pA to 10 nA for the InGaAs photo-diode and 10 pA to 100 nA for the Si photo-diode. An accurate photo-diode current measurement requires multiple samples. This was achieved by operating the pico-ammeter in sampling mode and by averaging over ten such samplings. Instrument drift during very long exposures was tracked by a reference photo-diode and subtracted from the photo-diode signal as shown in Fig. 5.2.

Our measurement technique requires that any deviation from photo-diode linearity

---

[1] The following band-pass filters were used: 700 nm central wavelength, 80 nm wide; 880 nm, 50 nm wide; 950 nm, 50 nm wide; and 1400 nm, 80 nm wide.

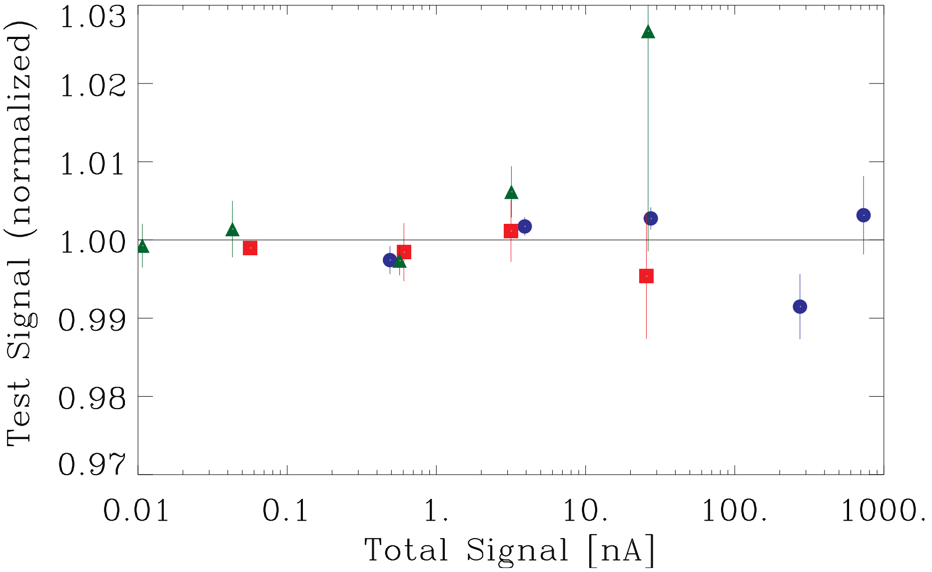


Figure 5.4: Normalized test signal as a function of total signal (base signal plus test signal) at approximately 9 pA (green triangles), 55 pA (red squares), and 488 pA (blue circles). A combined fit to all data results in a non-linearity of (0.08 ± 0.08)%/decade for the Si photo-diode. Note that the error bars on the normalized test signals represent mainly the systematic uncertainties in these measurements, since the statical uncertainties are negligible in comparison.

be well characterized and corrected for. Since precise linearity specifications were not available from the photo-diode vendors, photo-diode linearity was measured in our laboratory. We used a beam-addition method in which a small, constant "test signal" was intermittently added to "base signals" of varying intensities as illustrated in Fig. 5.3. A 70/30 beam splitter following the stabilized light source extracts a constant amount of light, the test signal, that is attenuated and guided through a shutter into the integrating sphere. The direct light beam, the base signal, passes through an aperture wheel allowing to vary base signal intensities. A photo-diode is mounted to the integrating sphere and, for different base signals, its response to the base signal alone and to base signal plus test signal is registered. The Si photo-diode, which served as the the primary monitoring photo-diode for the reciprocity measurement, was used for this calibration. It was illuminated[2] at different intensities spanning five orders of magnitude, and a power law model was fitted to evaluate the photo-diode linearity. In order to cover five orders of magnitude in illumination, three test signals of approximately 9 pA, 55 pA and 488 pA were used as shown in Fig. 5.4. The magnitudes of these test signals were fitted along with a power law exponent,

---

[2] Pass-band selected light of 950 ± 25 nm was used.

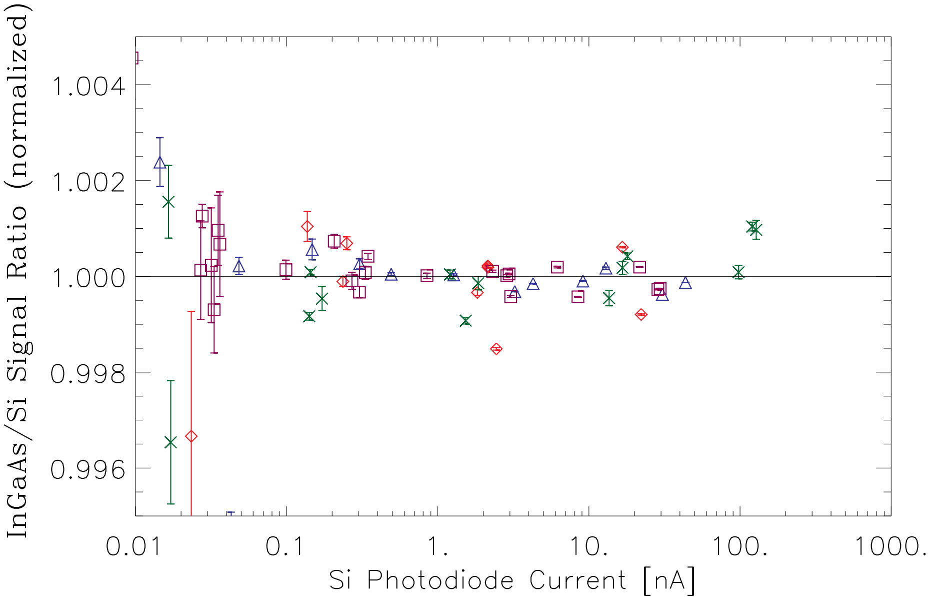


Figure 5.5: Normalized InGaAs photo-diode to Si photo-diode signal ratio as a function of Si photo-diode current. Measurements at wavelengths of 700 nm (red diamonds), 790 nm (green crosses), 880 nm (purple squares) and 950 nm (blue triangles) are shown.

resulting in a non-linearity of (0.08 ± 0.08)%/decade. This non-linearity was later utilized to correct the detector response measurements, and its error was assigned as a systematic uncertainty.

As shown in Fig. 5.5, the *relative* linearity of the Si and InGaAs photo-diodes is better than 0.1% over the dynamic range of illumination[3] and wavelength used during the reciprocity measurements. This agreement gives us confidence that the absolute linearity of the InGaAs photo-diode is also of the order of 0.1%/decade, which is consistent with previous linearity studies of Si and InGaAs photo-diodes (Budde 1979; Yoon et al. 2003).

5.1.3 Cryogenic System

Reciprocity failure in NIR devices was characterized at a baseline temperature of 140 K in an 8-inch dewar manufactured by IR Labs. The hold time of the system is typically 6 to 8 hours, longer than the longest sampling sequence which takes about 5 hours to complete. This guarantees that measurements are not disrupted by the liquid nitrogen refill process. For all measurements, the NIR detector was mounted

---

[3] The dynamic range corresponds to photo-diode currents between approximately 1 pA and 100 nA.

to a fixed copper heater plate which is weakly thermally coupled to the liquid nitrogen reservoir and thermally stabilized to 10 mK. The cool-down and warm-up ramp of 1 K/min as well as temperature stabilization of the NIR detectors at the operating temperature was controlled and monitored by a precision temperature controller (Lakeshore 330). With the temperature of the detector held constant at 140 K, the illumination system inside the dewar was allowed to cool down to below 200 K at the integrating sphere over a time period of about 8 hours. This is much colder than required to suppress thermal background radiation in the 1.7 μm detector material. A second temperature control loop was used to eliminate temperature dependence in the response of the two photo-diodes, which were always temperature stabilized at 270 K.[4] Additional measurements of the temperature dependence of reciprocity failure were performed. The temperature was set to 160 K, 120 K and 100 K for these.

5.1.4 Read-out and Control Electronics

For detector read-out and control, a commercially available data acquisition system from Astronomical Research Cameras (ARC) was used. In this system, 32 channels of parallel read-out are available from four 8-channel infrared video processor boards combined with clock driver boards and a 250 MHz timing and PCI card. This read-out electronics is described in detail in Leach & Low (2000a). Data are stored in FITS format for subsequent analysis. In the current set-up no shutter was employed and thus each detector pixel starts to integrate signal immediately after reset. Consequently, the shortest “illumination time” is determined by the amount of time it takes to read the array. In the default clocking mode (100 kHz) the read-out of the whole array takes 1.418 seconds. To reduce the illumination time, only a partial strip of the detector, 300×2048 pixels was read out for most of the measurements. This decreased the read-out time to 211 milliseconds. This readout mode is referred to as *stripe mode*. A fraction of the measurements was performed where the full detector was read out to probe possible spatial variation in reciprocity failure across a detector. This readout mode is referred to as *full mode*. The spatial resolution was sampled by subdividing the detector into tiles of 64×64 pixels in the *full mode* and 60×64 pixels in the *stripe mode*. This tiling reduces the uncertainty in the measurement due to photon shot noise and read noise.

Several detector characteristics depend on the bias voltage settings; the full integration capacity for instance is a function of the reset voltage. All measurements

---

[4] It was observed that at lower temperatures the InGaAs photo-diode response becomes slightly non-linear. See Chapter 6 for details.

reported here were performed with bias settings that were established to optimize low noise performance. The following voltages were applied: detector substrate voltage Dsub = 0.35 V, reset voltage Vreset = 0.10 V, pixel source follower bias voltage Vbiasgate = 2.45 V, and pixel source follower source voltage Vbiaspower = 3.23 V.

5.2 System Optimization

Many challenges had to be overcome to achieve the 0.1%/decade sensitivity to reciprocity failure in our system. Initial testing of the set-up indicated that it suffered from light leaks. The cryogenic ports identified as the source of the leaks were shielded, and the internal baffling system was extended to fully cover the detector to eliminate stray light in the system. The drifts in the photo-diode readout affecting low illumination measurements were first reduced with better cable shielding and grounding, and finally corrected for in the analysis using the signal from a reference photo-diode. It was noticed that dark images (where the aperture was closed) were brighter when the lamp was on than when it was off. This was caused by the light heating the aperture mounts causing them to glow in the NIR. It was mitigated by facing the reflective side of the mounts towards the light and by using a cold short-pass filter between the apertures and the integrating sphere for measurements below 1000 nm. At longer wavelengths, matched dark images were taken with the lamp on to allow a complete subtraction of this small dark glow. One of the greatest challenges involved the spectral mismatch of the detector and photo-diode responses. The comparison of the signals from both, the Si and InGaAs photo-diodes indicated that the pass-band filters leaked in the red. This was confirmed using a single wavelength laser. Either short-pass or long-pass filters were placed in the light path to improve out-of-band rejection. Monitoring photo-diode signal ratios also confirmed that the ND filters used were spectrally flat to better than 0.1% in the region we operated. This was not the case for other ND filters we checked. Using apertures instead of ND filters to control illumination avoids the spectral dependence issue. Hence they were used as the primary means of illumination control. It turned out, however, that the integrating sphere used was not large enough to fully wash out the image of the aperture at its entrance and therefore different apertures resulted in slightly different illumination patterns on the device. This was remedied by two layers of spectrally flat diffusers, added between the apertures and the integrating sphere. Ultimately, the different but complementary means of attenuating the illumination, the apertures and the ND filters, and the different spectral bands probed by the two photo-diodes were essential in reaching the required sensitivity in our measurements.

5.3 Test and Analysis Procedure

During a typical reciprocity measurement the detector was first reset then repeatedly read non-destructively in a procedure that is generally called Sample-Up-the-Ramp (SUR) mode, with up to 200 frames read during an exposure. For every SUR sequence "matched darks" were obtained. Measurement conditions for the matched darks were in every way identical to the reciprocity measurement conditions but exposures were taken with the aperture closed. The data sets obtained in this manner allow us to model and correct for integrated signal non-linearity.

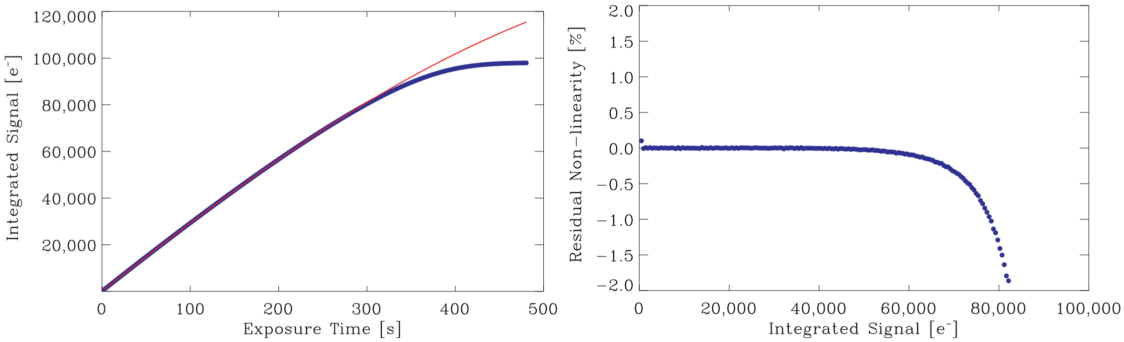


Figure 5.6: Modeling the integrated signal non-linearity in a HgCdTe detector. Left panel: Integrated signal in a HgCdTe detector as a function of time. The red curve is the result of the three-parameter fit of equation 5.4. Right panel: Deviation of the data from the fit versus integrated charge. The residual non-linearity is reduced to below 0.1% (1%) for signals below 60% (80%) of the saturation level.

5.3.1 Integrated Signal Non-Linearity

In order to properly evaluate detector response at differing illumination intensities, care must be taken to distinguish between reciprocity failure and integrated-signal non-linearity as the pixel integrates charge. The integrated signal, S, in the detector is parametrized as $S(t, F) = \int_0^t F(t') \times \epsilon(S) \mathrm{d}t'$, where $F(t)$ represents the detector count-rate as a function of time t, and $\epsilon(S)$ takes into account classical integrated signal non-linearity.

An ad-hoc three parameter model, intrinsically independent of the intensity level, was produced to describe the change in junction capacitance of the pixel as a function of integrated signal S. In a perfectly linear detector the voltage changes by a constant amount for each collected electron until the voltage is sufficient to forward-bias the detector diode. In a real detector this voltage change decreases with increasing S. Two parameters, a and b are used to parameterize this behavior, such that

$$\epsilon(S) = \frac{a + 1 - (a+1)^{\frac{S}{b}}}{a}\,, \tag{5.1}$$

where ϵ is defined to be unity when no charge, S, has been collected ($S = 0$), and zero when the pixel has "saturated" ($S = b$). The parameter a describes how quickly the junction capacitance is changing, as $a \to \infty$ the device becomes linear. The parameter b is the maximum voltage that the pixel can record, that is, the pixel saturation level. The rate of signal integration by the device can be written as

$$\frac{dS}{dt} = F(t)\epsilon(S)\,, \tag{5.2}$$

where $F(t)$ is the time dependent true flux.

Equation (5.2) can be integrated analyticaly only for certain models of the flux $F(t)$. We approximate the flux as constant illumination plus a dark current (with constant asymptotic value, d, and an exponentially decaying component, d_e). The flux can then be written as

$$F(t) = F_0 + d + d_e \cdot exp(-\frac{t}{\tau})\,. \tag{5.3}$$

The dark current is fitted separately with the exponentially decaying model using data sets obtained in the dark resulting in the values of d, d_e, and τ being known at the time of the integrated signal fit.

Equation (5.2) is then integrated to the form

$$S(t) = \frac{b}{\log(1+a)} \log\left(\frac{1+a}{1+\exp\left(\frac{\alpha}{b} + \frac{\alpha}{ab} + \frac{\beta}{b}\right)}\right)\,, \tag{5.4}$$

with α and β defined as

$$\alpha = \left(d + F_0 - dt - F_0 t + d_e\tau(e^{-\frac{t}{\tau}} - e^{-\frac{1}{\tau}})\right)\log(1+a)\,, \tag{5.5}$$

$$\beta = \left(\left(d + F_0 + d_e\tau(1 - e^{-\frac{1}{\tau}})\right)(-1 - \frac{1}{a}) + \frac{b\log a}{\log(1+a)}\right)\log(1+a)\,. \tag{5.6}$$

After discarding the first frame to avoid turn-on effects, each i^{th} SUR image, $S(t_i) -$

$S(t_{i-1})$, is fitted for the three parameters, a, b and F_0. The value of F_0 serves as the detector response independent of the integrated-signal non-linearity and is divided by the corresponding photo-diode current to compute the normalized flux ratio.

As can be seen in Fig. 5.6, applying the parametrization from equation 5.4 describes the observed behavior well (see left panel of figure). After the correction, the integrated-signal non-linearity for signals below 60% of the saturation level is less than 0.1% (see right panel of figure). Exposed images and matched darks were included in the fit procedure used in calculating the NIR detector response F_0 for various illumination levels. In our ad-hoc model, the two non-linearity parameters were fitted simultaneously to all the different illumination intensity sets, while the flux was fit separately. This ensures that reciprocity failure is not hidden in the possible degeneracy of those parameters. It also reduces the uncertainties on the estimated parameters. As a check we also fitted each illumination set separately. The values for reciprocity failure so obtained agreed with the combined fit results. In addition, an analysis was performed without accounting for integrated signal non-linearity while keeping the total integrated signal roughly constant. This analysis is discussed in §5.4.1.1.

5.3.2 Flux Normalization

Monitoring photo-diode currents were recorded for each frame in the sample. Long exposures over several hours were typical at the lowest illumination levels of a few electrons/pixel/second at the detector. It was observed that at the most sensitive setting the pico-ammeter drifts at the 10% level. Those fluctuations were tracked by a reference photo-diode connected to a pico-ammeter and removed from the data as shown in Fig. 5.2. The residual variation in the current measurement is dominated by statistical fluctuations and the variance of the mean improves linearly with the number of measurements in the exposure. The Si photo-diode itself was found to deviate from linearity at a level of (0.08 ± 0.08)%/decade, requiring a correction that reduced the photo-diode signal by this amount. The uncertainty in the Si photo-diode calibration along with the InGaAs to Si photo-diode ratios constitute the systematic limit of our sensitivity to reciprocity failure of 0.1%/decade. The fitted detector response is divided by the photo-diode current resulting in the flux ratios shown in Fig. 5.7. Normalized flux ratios were obtained at different illumination intensities and at different wavelengths. At wavelengths below 1000 nm, current readings from the Si photo-diode and above 1000 nm, readings from the InGaAs photo-diode were used for calculating the flux ratios.

5.4 Measurements and Results

Detector characterization was initially performed at a single temperature (140 K) followed by measurements at several wavelengths to test a possible wavelength dependence of reciprocity failure. In later measurements the temperature was also varied to investigate temperature dependence. First the results from measurements on each of the four detectors at 140 K are discussed. In §5.4.4 it is shown how reciprocity failure can be mitigated by lowering the device temperature. An overview of the measurements at the baseline temperature of 140 K can be found in Table 5.1. Note that the quoted uncertainties are statistical only and do not include the overall 0.1% systematic uncertainty. The impact of reciprocity failure is briefly discussed in §5.4.5 and its possible dependence on the exposure time is addressed in §5.4.6.

Detector	Wavelength [nm]	Reciprocity Failure [% decade^{-1}] *Stripe Mode*	*Full Mode*
H2RG-102	700	0.35 ± 0.04	
	790	0.35 ± 0.03	
	880	0.36 ± 0.05	
	950	0.29 ± 0.04	
	1400	0.38 ± 0.05	
H2RG-142	790	0.38 ± 0.03	0.53 ± 0.14
	950	0.48 ± 0.07	
	1400	0.33 ± 0.04	
H2RG-236	790	10.9 ± 0.5	10.3 ± 0.6
	950	11.9 ± 0.5	
	1400	11.7 ± 0.5	10.6 ± 1.9
H2RG-238	790	5.1 ± 0.7[a]	4.0 ± 0.8
	950	4.4 ± 0.4	

Table 5.1: Reciprocity failure data at 140 K.
[a] *Full* mode data analyzed as *stripe* mode.

5.4.1 H2RG-102 at 140 K

Device H2RG-102 was manufactured early on during the *SNAP/Joint Dark Energy Mission* (*JDEM*) R&D program and was delivered in 2005. The QE is greater than 90% from 0.9 μm to 1.7 μm and about 40% at 0.45 μm. The dark current and read noise performance is very good; the Fowler-1 noise is 25 e^-. Unlike devices produced later, this detector is mounted on a molybdenum pedestal. The multiplexer is of type HAWAII-2RG-A0. This detector exhibits low reciprocity failure (0.35 ± 0.03) % decade^{-1} at 790 nm and shows no wavelength dependence. Figure 5.7 shows the flux ratios as a function of count rate with a logarithmic fit (linear in log illumination) that describes the data well. As indicated in the figure, reciprocity failure for the H2RG-102 detector tested in our set-up is very low. The NIR detector count rate is scaled relative to the photo-diode current to remove flux dependence from the horizontal axis. Measurements were performed at five different wavelengths (700 nm, 790 nm, 880 nm, 950 nm and 1400 nm) with no significant wavelength dependence observed as shown in Fig. 5.8. Measured values for the reciprocity failure at the five wavelengths (in %/decade) are 0.35 ± 0.04, 0.35 ± 0.03, 0.36 ± 0.04, 0.29 ± 0.04, and 0.38 ± 0.05. These reciprocity failure values are subject to a 0.08%/decade systematic uncertainty in the photo-diode non-linearity correction. This result contrasts with the strong wavelength dependence for reciprocity failure in all three NICMOS detectors.

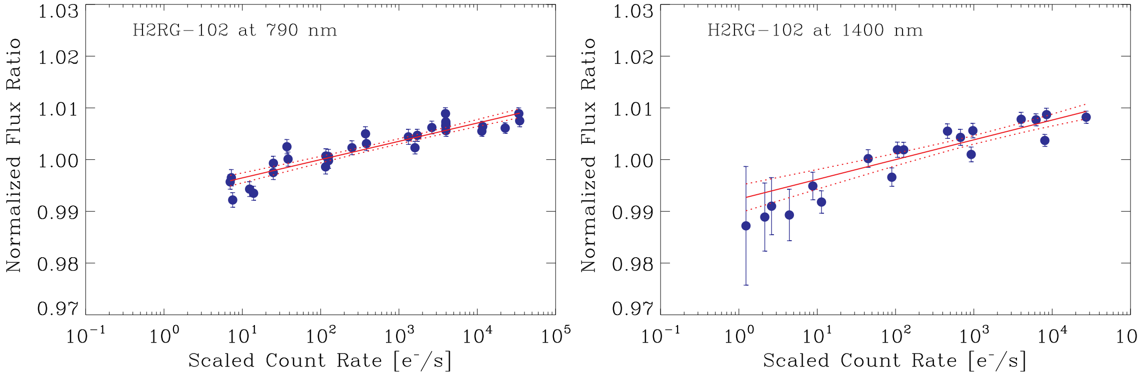


Figure 5.7: Reciprocity failure versus scaled count rate in device H2RG-102 at 790 nm (upper panel) and 1400 nm (lower panel). The solid lines indicate a logarithmic fit to the data points. The 1σ error bands (dotted lines) include the point-to-point statistical and systematic uncertainties, but not the systematic uncertainty due the photo-diode calibration of 0.08%/decade. The measured values for the reciprocity failure at 790 nm is $(0.35 \pm 0.03\,(\text{stat.}) \pm 0.08\,(\text{syst.}))$%/decade, and $(0.38 \pm 0.05\,(\text{stat.}) \pm 0.08\,(\text{syst.}))$%/decade at 1400 nm.

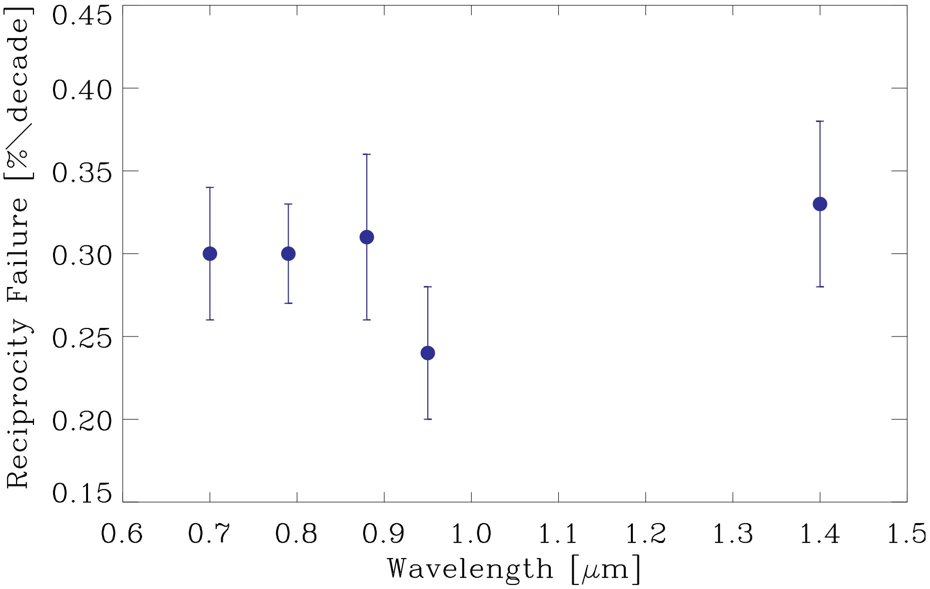


Figure 5.8: Reciprocity failure as a function of wavelength for device H2RG-102.

5.4.1.1 Model-Independent Analysis

In addition to the modeled signal non-linearity correction above, a model-independent signal integration method was utilized to confirm that the fitting procedure described in §5.3.1 did not bias the reciprocity failure measurements. This was accomplished by utilizing the SUR acquisition mode. A frame can be selected for each data sequence where the integrated signal is within ± 5 % of a target value which means that the integrated signal non-linearity does not affect our results and a ratio of average detector flux to photo-diode flux is used directly. Figure 5.9 shows the results of the signal integration approach to reciprocity failure characterization at 790 nm and 1400 nm. It somewhat limits the dynamic range of observations since the integration times at low intensities needed to reach a certain total signal are far too long. The values obtained using this method were $0.35^{+0.01}_{-0.02}$%/decade at 790 nm and $0.37^{+0.01}_{-0.05}$%/decade at 1400 nm. These results are comparable to the what was obtained with modeled signal non-linearity correction. As can be seen in Figure 5.9, the individual illumination intensity samples suffer from large uncertainties due to first-frame readout noise. In order to extend the dynamic range of observation and decrease the measurement uncertainties, the final magnitudes of reciprocity failure are obtained by modeling and correcting for signal non-linearity, as in Figure 5.7.

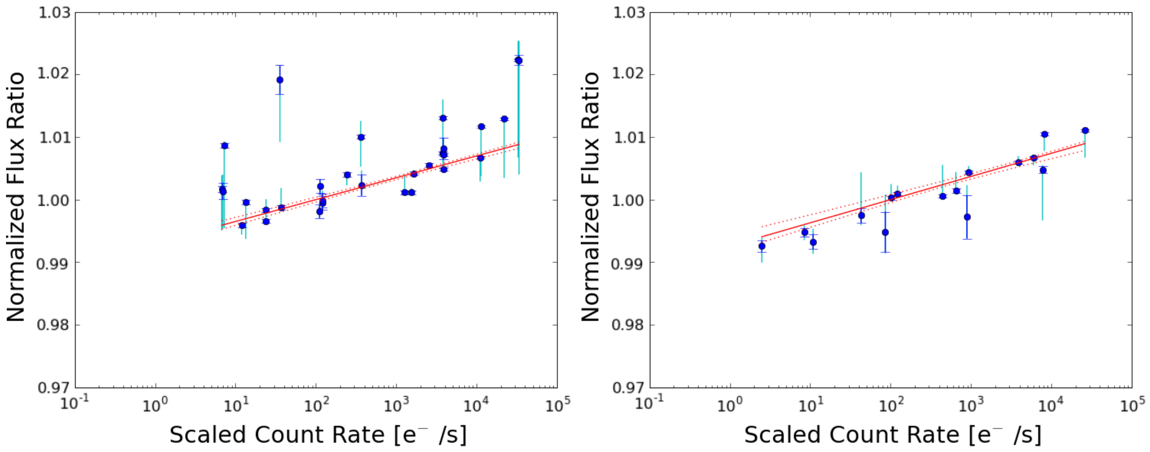


Figure 5.9: Reciprocity failure versus scaled count rate in device H2RG-102 at 790 nm (left panel) and 1400 nm (right panel) derived in a model-independent way. The dark blue error bars show the statistical uncertainties for each measurement. The cyan error bars show the systematic uncertainties that are dominated by first-frame readout errors. The solid lines indicate a logarithmic fit to the data points. The 1σ error bands (dotted lines) include the point-to-point statistical and systematic uncertainties, but not the systematic uncertainty due the photo-diode calibration of 0.08%/decade.

5.4.2 H2RG-142 at 140 K

Device H2RG-142 came from the fifth manufacturing run for *SNAP*. It was mounted on a SiC pedestal specifically developed for *SNAP*/*JDEM* to provide a good thermal match to the multiplexer. Devices from this run were also mated to the HAWAII-2RG-A0 multiplexer. H2RG-142 has high QE and low read noise. It exhibits a somewhat larger number of hot pixels than H2RG-102 but is otherwise cosmetically good. Figure 5.10 shows reciprocity failure of $(0.38 \pm 0.03)\,\%\,\text{decade}^{-1}$ at 790 nm in *stripe mode*. The average reciprocity failure value measured for this device was very similar to detector H2RG-102 at all wavelengths. In addition to the *stripe mode* measurements the structure of reciprocity failure was also characterized in the *full mode*. Although the signal to noise ratio was low, non-linearity variations in the detector did appear in a range from 0.35 to $0.85\,\%\,\text{decade}^{-1}$. In particular one corner of the device exhibited larger reciprocity failure. This map is not shown however a more impressive example is discussed in §5.4.3.

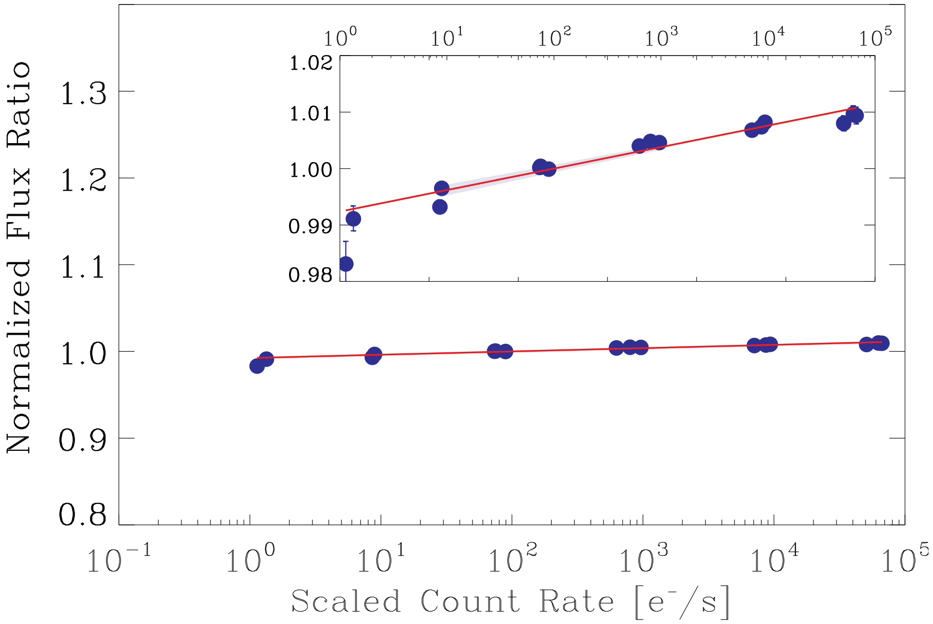


Figure 5.10: Reciprocity failure measured in *stripe mode* for device H2RG-142 at 790 nm. The ordinate scale was set to allow a direct comparison with detectors H2RG-236 and H2RG-238. A magnified scale is shown in the insert. The 68 % confidence level is indicated by the shaded area.

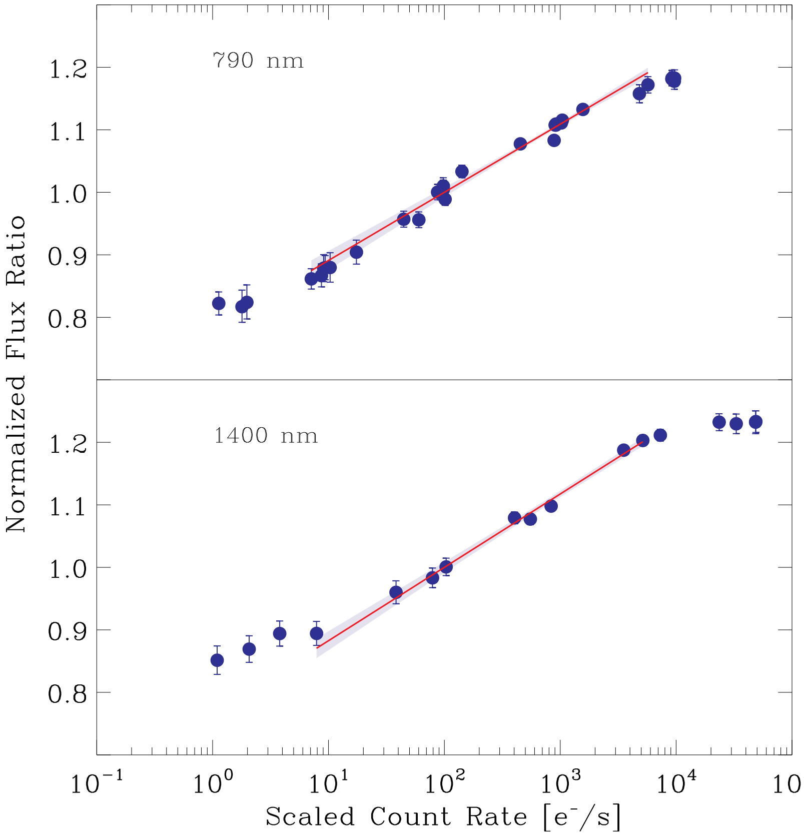


Figure 5.11: Average reciprocity failure measured in device H2RG-236 at 790 nm (top panel) and 1400 nm (bottom panel). Data was taken in the *stripe mode*. The 68 % confidence level is indicated by the shaded area.

5.4.3 H2RG-236 and H2RG-238 at 140 K

Devices H2RG-236 and H2RG-238 were produced during the sixth manufacturing run of the *SNAP / JDEM* R&D program. Like device H2RG-142 they both are mounted on a SiC pedestal but unlike that device, they were hybridized to a newer multiplexer, the HAWAII-2RG-A1 designed in part to reduce capacitive coupling between neighboring pixels (Brown et al. 2006). Both devices have low dark current and read noise and are very good cosmetically. Quantum efficiency of both devices is lower than in earlier detectors but is exceptionally uniform when measured at high flux. The average reciprocity failure measured in *stripe mode* for device H2RG-236, shown in Figure 5.11, is $(10.9 \pm 0.5)\,\%\,\mathrm{decade}^{-1}$ and $(11.7 \pm 0.5)\,\%\,\mathrm{decade}^{-1}$ at 790 nm and 1400 nm, respectively. The results from the two measurements are very similar, emphasizing the insensitivity of reciprocity failure to the wavelength of the illumination for these detectors. Data taken at 1400 nm and 950 nm (not shown in Figure 5.11) revealed that a linear fit is only representative for illumination levels between roughly 10 counts s^{-1} and 10,000 counts s^{-1}. Outside this range the detector response appears to become linear, indicating a saturation effect at high illumination levels and possibly a turn-on threshold at low illumination levels. Detector H2RG-236 showed the largest reciprocity failure of the four devices measured. As a check, a model-independent approach to this measurement was used and returned consistent results as in §5.4.1.1 further indicating that the integrated signal non-linearity modeling discussed in §5.3.1 does not bias our results.

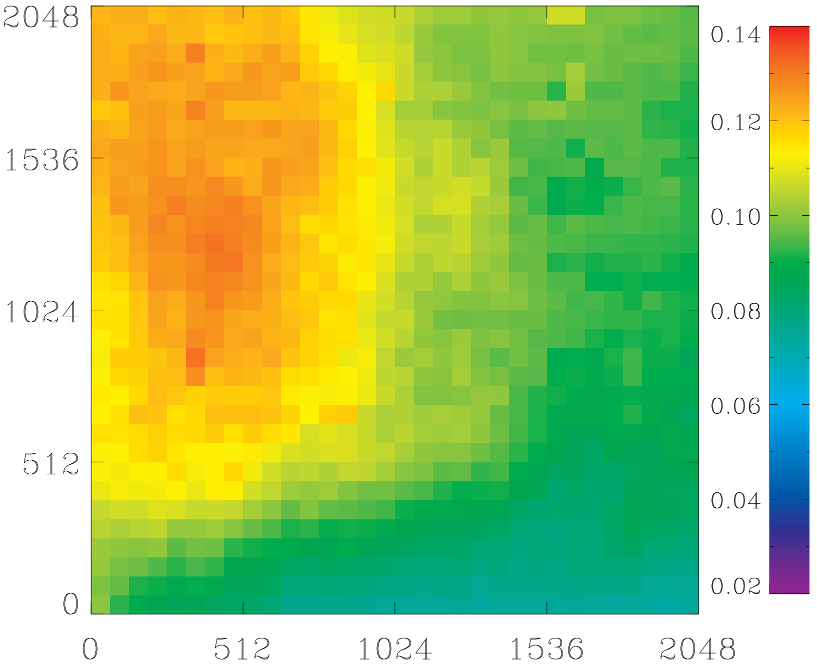


Figure 5.12: Reciprocity failure map for device H2RG-236 at 790 nm. The scale is in $\%\,\mathrm{decade}^{-1}$.

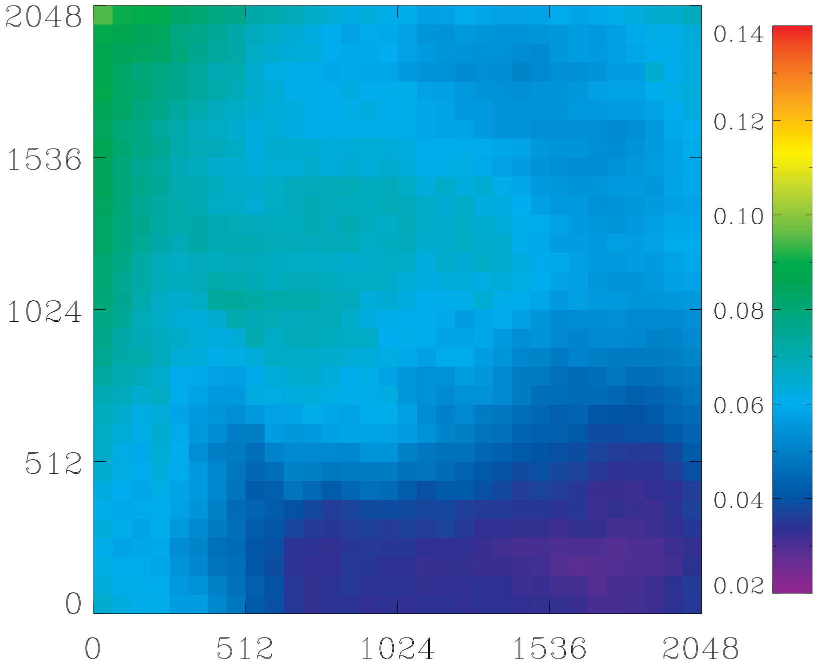


Figure 5.13: Reciprocity failure map for device H2RG-238 at 950 nm. The scale is in % decade^{-1}.

Strong spatial variation of reciprocity failure was observed in these two devices, ranging from 7.3 % decade^{-1} to 13.1 % decade^{-1} for device H2RG-236 (Figure 5.12) and from 2.9 % decade^{-1} to 9.5 % decade^{-1} for device H2RG-238 (Figure 5.13). It is worth noting that for such a device, simply correcting for the average reciprocity failure without accounting for spatial structure will result in a large residual uncertainty in photometric measurements.

5.4.4 Temperature Dependence

In an attempt to better understand the physical mechanisms that lead to reciprocity failure, it was investigated how reciprocity failure is affected by device temperature. Detectors H2RG-142 and H2RG-236, low and high reciprocity devices, respectively, were tested at temperatures ranging from 100 K to 160 K. These tests revealed that flux dependent non-linearity can be "frozen out" at sufficiently low temperatures. The results from the two detectors, shown in Table 5.2, suggest that this freeze-out temperature depends on the amount of reciprocity failure in a particular detector and will therefore vary for different detectors.

5.4.5 Reciprocity Failure and QE

For detectors that exhibit reciprocity failure, care must be taken when measuring quantum efficiency. Reciprocity failure will bias QE measurements towards higher

Temperature	Reciprocity Failure [% decade^{-1}]	
	H2RG-142	H2RG-236
160	2.2 ± 0.3	
150		10.9 ± 0.9
140	0.48 ± 0.07	11.9 ± 0.5
120	0.15 ± 0.07	3.0 ± 0.7
100		0.1 ± 0.4

Table 5.2: Reciprocity failure versus temperature. Data obtained in *stripe* mode at 950nm.

values at high illumination levels and towards lower QE values at low illumination levels. In addition, spatial nonuniformity of reciprocity failure across a detector will alter the apparent device uniformity as a function of the illumination intensity. One possible approach is to measure QE at sufficiently low temperature to suppress reciprocity failure in order to reveal the "true" QE. This topic, among others, is further explored in Chapter 6.

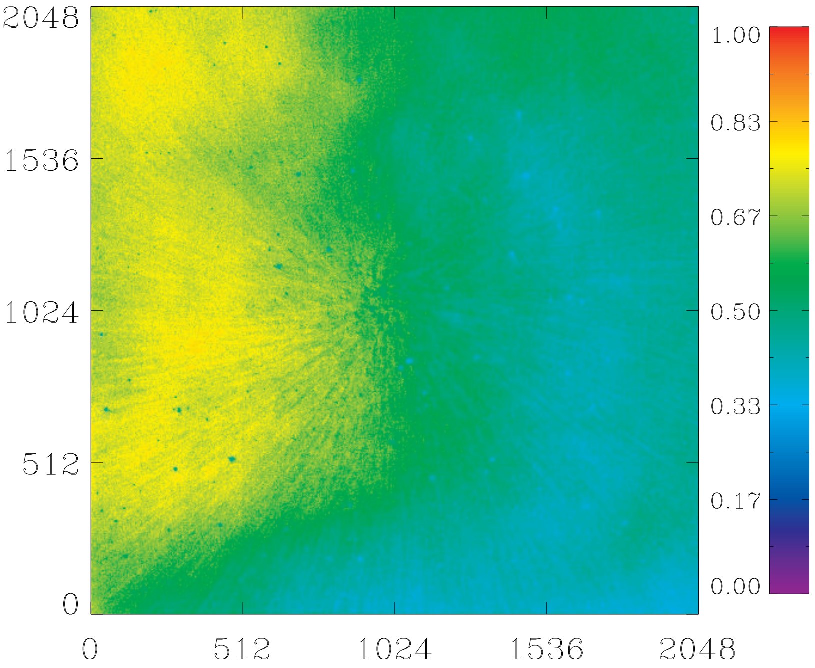


Figure 5.14: Ratio of two H2RG-236 flat field images with a factor of 1000 difference in flux. The observed large scale structure is due to reciprocity failure. The measurement was performed at 140 K.

Precise characterization of reciprocity failure is a rather elaborate procedure and requires a specialized experimental setup. However, a simple measurement can reveal possible *spatial structure* in a detector's reciprocity failure. Using a standard flat field illumination test setup two flat field images were produced, one at a very high illumination intensity and a second at a very low illumination intensity. The ratio of these two images, shown in Figure 5.14, displays the same spatial variability as the reciprocity failure map for this detector shown in Figure 5.12. Such a measurement may therefore be used as a simple test that does not require any special equipment beyond a basic illumination system. However, some caveats apply. This test will only reveal spatial structure in reciprocity failure of a device, and will not produce an absolute value for the strength of reciprocity failure, nor will it reveal reciprocity failure in detectors where the effect is spatially uniform. An example of this approach being used was a measurement performed by our collaborators (Smith 2010) for device H2RG-220 which put a lower limit on its reciprocity failure of 1.2% decade^{-1} as shown in Table 5.3.

5.4.6 Reciprocity Failure at Constant Exposure Time

Reciprocity failure was measured by integrating charge to a constant level for all but the lowest count rates (less than about 3 electrons/second) while adjusting the flux. However, it must be noted that the count rate that we characterize reciprocity failure against is degenerate with the exposure time when the integrated signal level is held constant. Hence, one could argue that reciprocity failure may be characterized instead as a function of exposure time. In order to break this degeneracy we re-analyzed the 790 nm data for device H2RG-236 in a mode where we hold the exposure time, and not the integrated signal, constant. This analysis relies heavily on our modeling of the integrated signal non-linearity in §5.3.1 since the device response fits are cut off at varying integration levels. Nevertheless, this analysis can give us an idea of the importance of exposure time which may shed light on possible reciprocity failure causes like charge trapping. Figure 5.15 shows the reciprocity failure fits at nine different exposure times (labeled in the panels). The dynamic range for each exposure time is low but a fit can be made nevertheless. $1\,\sigma$ uncertainty bands around the best fit value are shown as shaded red regions. The reciprocity failure obtained in the regular fit in the top panel of Figure 5.11 is indicated by a solid black line in each panel. While the best fit values vary at different exposure times, they are all consistent with the regular model of reciprocity failure, usually within the $1\,\sigma$ bounds. This indicates that reciprocity failure is truly a function of the count rate and does not depend on time.

This casts doubt on theories in which reciprocity failure is caused by a charge trapping mechanism. In such scenario charges would be trapped with a characteristic time constant. The density of long term traps would have to be substantially higher

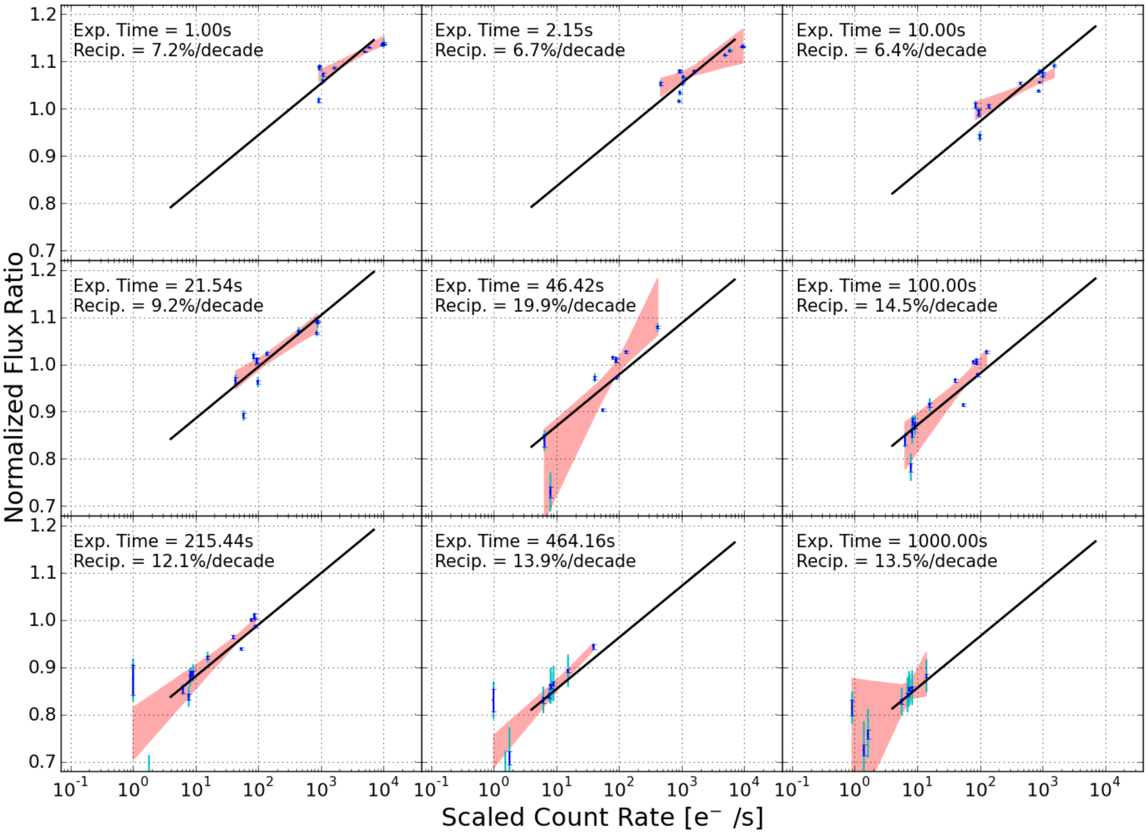
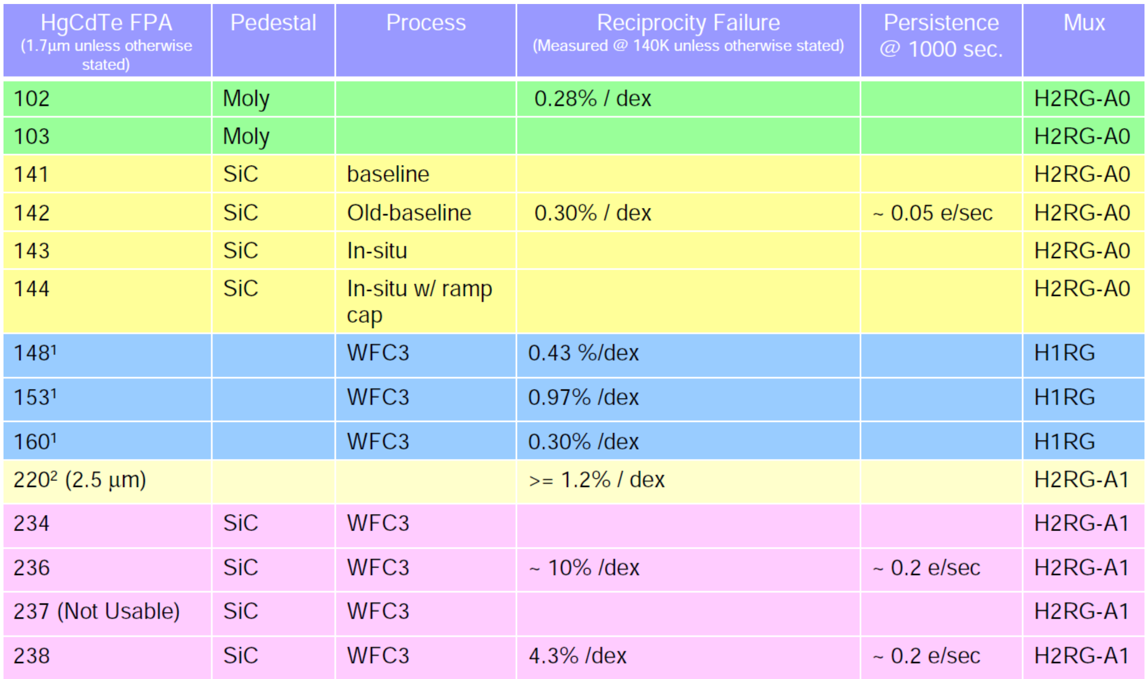


Figure 5.15: Reciprocity Failure at 9 different exposure times for H2RG-236. The blue lines show the statistical uncertainties around the normalized (at 100 electrons/second) flux ratios. Cyan error bars show the total uncertainties (systematic and statistical added in quadrature) of each ratio if larger than the statistical ones. Shaded red regions shows the $1\,\sigma$ bounds around the best fit reciprocity model for each exposure time. In addition, the default reciprocity failure from the top panel of Figure 5.11 is shown as a solid black line shifted to the appropriate intercept. The exposure time and fitted reciprocity failure values are indicated in each panel.

than the density of short term traps since more electrons appear to be lost at low illumination. This seems unlikely as one would naively expect the density of traps to be inversely proportional to their time constant. Nevertheless, in this test where charge at different illumination intensities is integrated over a constant exposure time the same traps should have been activated leading to a constant response when plotted vs the illumination intensity. A small *negative* slope may in fact have been possible since at high flux the depletion region in each pixel diode would have been filled more, potentially exposing more traps. However, the slopes of reciprocity failure in Figure 5.15 all appear consistent with the overall measured value.

5.5 Comparison To Other Measurements



HgCdTe FPA (1.7µm unless otherwise stated)	Pedestal	Process	Reciprocity Failure (Measured @ 140K unless otherwise stated)	Persistence @ 1000 sec.	Mux
102	Moly		0.28% / dex		H2RG-A0
103	Moly				H2RG-A0
141	SiC	baseline			H2RG-A0
142	SiC	Old-baseline	0.30% / dex	~ 0.05 e/sec	H2RG-A0
143	SiC	In-situ			H2RG-A0
144	SiC	In-situ w/ ramp cap			H2RG-A0
148[1]		WFC3	0.43 %/dex		H1RG
153[1]		WFC3	0.97% /dex		H1RG
160[1]		WFC3	0.30% /dex		H1RG
220[2] (2.5 µm)			>= 1.2% / dex		H2RG-A1
234	SiC	WFC3			H2RG-A1
236	SiC	WFC3	~ 10% /dex	~ 0.2 e/sec	H2RG-A1
237 (Not Usable)	SiC	WFC3			H2RG-A1
238	SiC	WFC3	4.3% /dex	~ 0.2 e/sec	H2RG-A1

Table 5.3: Reciprocity failure of various HgCdTe detectors. Devices 102, 142, 236 and 238 have been measured at 140 K. Devices 148, 153, 160 are likely to have been measured at 145 K and device 220 is likely to have been measured at a temperature below 140 K. Additional data from: [1]Hill et al. (2010) and [2]Smith (2010)

Table 5.3 contains a summary of reciprocity failure measurements for various HgCdTe devices developed by TIS for *SNAP* and WFC3. Note that only few of the devices listed have had reciprocity failure measured. It also contains some additional information like the base pedestal the detector is mounted on, the manufacturing process, long term persistence and multiplexer type. A trend is visible where the 100-series devices have reciprocity failure less than 1%/decade while the 200-series detectors have larger reciprocity failure values. This may be related to the multiplexer design. Insufficient data is available to make any definite statements.

5.6 Possible Causes

Although a detector's reciprocity failure can be large, it will likely be possible to correct for it. If a sufficient amount of calibration data is obtained it should be possible to correct for reciprocity failure on a pixel by pixel level. Cooling detectors that exhibit strong reciprocity failure provides a straightforward mitigation strategy although the required temperature may vary for individual devices. While, for example, detector H2RG-142 will likely not exhibit noticeable reciprocity failure at 120 K, device H2RG-236 would have to be cooled below 100 K. Nevertheless, if the root cause of reciprocity failure could be identified it may be possible to eliminate it all together so that the costly and time consuming calibration process can be avoided. This would also be desirable if reciprocity failure changes as a function of time due to, for example, radiation damage. We therefore attempt to correlate reciprocity failure pattern with other know detector response structures as well as explore to a limited degree the effects of the readout multiplexer choice. Finally, we speculate about other effects that we could not test for.

5.6.1 Response Structure Correlations

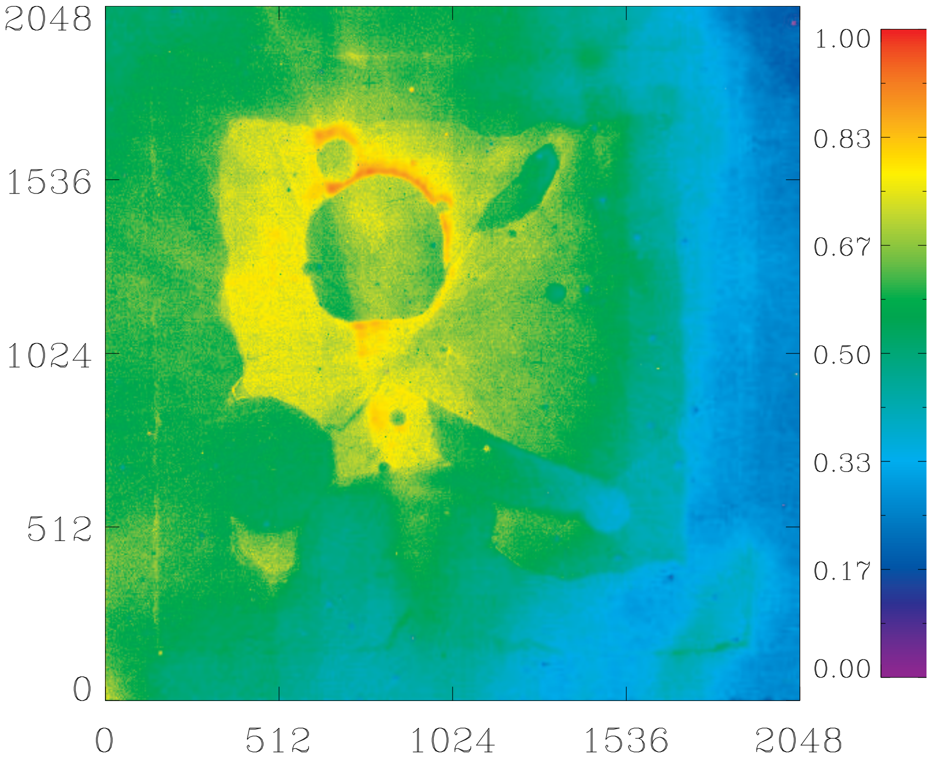


Figure 5.16: Image persistence in H2RG-236.

The observed spatial nonuniformity in reciprocity failure provides an opportunity to investigate a possible correlation with other detector properties such as dark current, QE near cutoff [5], and image persistence. Therefore the cross correlation between

---

[5] In all HgCdTe devices that were tested, strong QE variations are observed near the cutoff wavelength. This is caused by inconsistencies in the doping of the HgCdTe material by the MBE process.

the spatial structure of reciprocity failure and the other properties was computed for device H2RG-236. This particular detector was selected because of the pronounced spatial nonuniformity in reciprocity failure. The correlation coefficient was normalized to have a value between -1 and 1 for fully anti-correlated structure and identical structure, respectively. A value of zero represents the absence of correlation. Both, the 790 nm and the 1400 nm reciprocity failure data were used for this analysis as shown in Table 5.4. The correlation coefficient for the reciprocity failure maps at those two wavelengths is 0.92, indicating that not only the average reciprocity is independent of wavelength but also the spatial structure. The image persistence of device H2RG-236 (Figure 5.16) has the largest correlation coefficient due to similar large-scale structures. However, prominent features seen there do not appear in the flat-field ratio map (Figure 5.12). Therefore, image persistence for this device cannot be conclusively linked to reciprocity failure.

	790 nm	1400 nm
Dark Current	-0.41	-0.42
QE (1750 nm)	0.11	0.15
Persistence	0.70	0.57
Conversion Gain	-0.09	0.00

Table 5.4: Correlation of reciprocity failure and other detector properties at 790 nm and 1400 nm.

5.6.2 Multiplexer Choice

The two detectors that show low reciprocity failure, H2RG-102 and H2RG-142, and the two detectors that show high reciprocity failure, H2RG-236 and H2RG-238, differ in the type of multiplexer used for device readout. The 100-series detectors were hybridized to the HAWAII-2RG-A0 multiplexer while for the 200-series the redesigned HAWAII-2RG-A1 multiplexer was used. It was investigated whether the change in the multiplexer design was responsible for the large discrepancy in reciprocity failure between the 100 and 200 series. For this test an external RC circuit with a large capacitance and a precisely measured selectable resistance was used. The RC circuit was charged, simulating charge collecting at the pixel node, and read out by the multiplexer. Using the RC circuit instead of the detector diode allowed to measure the linearity response of the multiplexer by varying the circuit's impedance. The test was performed with the multiplexers of devices H2RG-142 and H2RG-238. No difference in multiplexer voltage readout linearity was observed, indicating that the difference in multiplexer readout electronics alone is not responsible for the observed difference in reciprocity failure.

5.6.3 Other Possible Causes?

At present the fundamental mechanism that leads to reciprocity failure is not understood. The comparison of spatial structures in characteristic maps discussed above does not provide a satisfactory suggestion of correlation between reciprocity failure and any other detector characteristic. In fact the only correlations that has been observed apply to the image persistence and multiplexer type. Note that this observation is based on the very limited sample of detectors discussed here and may not be a general property of HgCdTe detectors. Small leakage currents due to Ohmic parasitic resistance across the integrating field effect transistor can be excluded as cause for reciprocity failure because they would not reproduce the observed power-law behavior. However, non-linear leakage currents, typical for diodes, may provide an explanation for this effect. Furthermore, a charge trapping mechanism has been suggested as the underlying mechanism for image persistence (Smith et al. 2008), and it is conceivable that such a process also accounts for reciprocity failure though one may expect this mechanism to be a function of exposure time instead of flux and hence incompatible with our observations in §5.4.6.

5.7 Summary

Reciprocity failure was measured in four devices developed as part of the *SNAP/JDEM* R&D program with an overall sensitivity of 0.1% per decade in illumination intensity. It was found to vary from device to device with detector-averaged values (in $\%\,\mathrm{decade}^{-1}$ at 790 nm) of 0.35 ± 0.03 for H2RG-102, 0.38 ± 0.03 for H2RG-142, 10.9 ± 0.5 for H2RG-236 and 5.1 ± 0.7 for H2RG-236 with an overall 0.08 systematic uncertatiny. In addition, spatial variation of reciprocity failure was observed in all three devices that were tested in the *full* readout mode. A wavelength dependence, such as reported for the NICMOS detectors on *HST*, was not observed. The fabrication of *JDEM/SNAP* devices is based on WFC3 detector development. This is reflected in measurements on the final candidate detectors for WFC3 which show very similar results as H2RG-102 and H2RG-142 (Hill et al. 2009). The WFC3 team reports reciprocity failure ranging from 0.3%/decade to 0.97%/decade for three detectors. As they point out, this is significantly smaller than the effect seen for the 2.5 μm HgCdTe NICMOS detectors on *HST* (6%/decade). However, reciprocity failure observed in H2RG-236 and H2RG-238 was significantly larger, comparable to the NICMOS measured values.

Reciprocity failure causes a systematic error in measurements of faint astronomical sources relative to bright standards. If not corrected for, an observation spanning three decades in illumination could suffer from a 1% (in low reciprocity devices) to 30% (in high reciprocity devices) error in the flux determination. Such a device would, if used for supernova cosmology for example, lead to an incorrect overestimate of the acceleration of the universe. In addition, this non-linearity has to be accounted for when performing a standard detector characterization such as measuring QE. The value of QE and its spatial uniformity depends on the intensity of the light at which they are measured.

Because of the wide range of reciprocity failure from one detector to another and of its spatial structure, reciprocity failure calibration presents a challenge. Furthermore, it is currently unknown if on-orbit radiation damage may alter it. Without a fundamental understanding of the underlying mechanism, reciprocity failure is therefore best addressed by the selection of “low reciprocity failure” devices and by cooling them sufficiently.

CHAPTER 6

Beyond Quantum Efficiency: A Comprehensive NIR Detector Response Study

Future space missions such as Euclid (Amiaux et al. 2012) and WFIRST (Spergel et al. 2013) plan on utilizing Mercury-Cadmium-Telluride (HgCdTe) near-infrared (NIR) detector technology which provides access to high redshift objects and structures in our universe. The performance of HgCdTe NIR detectors (Norton 2002) has been significantly improved during the last two decades and the widespread adoption of this technology by the astronomical community led to major characterization and testing efforts. Laboratory studies and use on telescopes have revealed detailed performance features and have contributed to a better understanding of those devices (e.g. Finger et al. 2004; McCullough et al. 2008; Moore 2006; Barron et al. 2007; Biesiadzinski et al. 2011a). We extended the efforts discussed in Chapter 5 to study the quantum efficiency (QE) of such devices in detail. This chapter is largely a reproduction of Biesiadzinski et al. (2013) in preparation to be published.

High QE with well characterized response at the pixel level greatly enhances the quality of astronomical detectors as it impacts survey speed and photometric precision. Measurements of QE are typically performed by illuminating the detector with a light source and comparing the signal observed in the detector with the signal measured in a well calibrated photo-sensor. Narrow band filters or monochromators permit the study of QE as a function of wavelength and uniform illumination needs to be produced for a typical detector area of several cm^2 (e.g. Schubnell et al. 2009).

As part of a program to characterize NIR detectors for the former Joint Dark Energy Mission (JDEM) (Schubnell et al. 2006) we performed a detailed study of the photon to electron conversion process in a HAWAII-2RG HgCdTe NIR detector (H2RG-236), manufactured by Teledyne Imaging Sensors (TIS). During the course of this study we have come to understand that detector QE cannot be adequately characterized without taking into account non-linear effects. For example, reciprocity failure (e.g. Biesiadzinski et al. 2011a) results in a larger measured QE at high flux levels than at low flux levels for an identical integrated photon count.

Other, typically less pronounced effects, such as non-linearities in both the detector and reference photodiodes, uncertainty associated with the conversion gain measurement and capacitive coupling (Brown et al. 2006) will be reflected in the measured detector QE. All those effects may depend on temperature and bias voltage settings. Furthermore, spatial variations in the effective band gap will result in local variations in the QE near the detector cut-off wavelength.

Here we describe an experimental set-up specifically tuned to measure the effects listed above on detector QE and we discuss their impact on the precision of the QE determination. Throughout all measurements we carefully accounted for statistical and systematic uncertainties (see §6.3.1). We have determined that we controlled *relative* systematic uncertainties to 3% for wavelengths longer than 800 nm. For the absolute QE measurement a 3.4% uncertainty due to conversion gain estimates discussed in §6.2.5 has to be added.

6.1 Setup

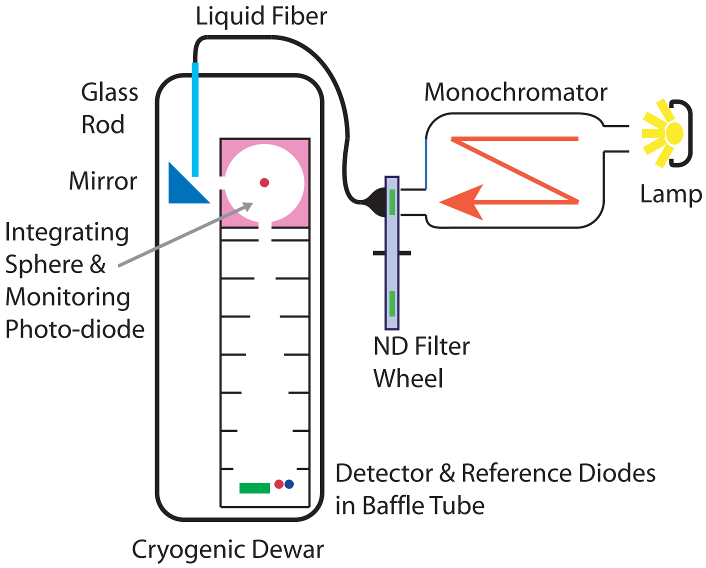


Figure 6.1: Main components of the QE experimental setup. See text for details.

A schematic representation of the setup used for the measurements described here is shown in Figure 6.1. The system produces well defined narrow-band light between 400 nm and 1800 nm uniformly illuminating a HAWAII-2RG (H2RG) NIR detector and two reference diodes (Figure 6.2). The detector quantum efficiency is determined by comparing the photon flux derived from the diode currents with the photon flux recorded by the NIR detector.

6.1.1 Illumination

An incandescent 200 W Quartz-Tungsten-Halogen (QTH) lamp directs light into a monochromator equipped with two selectable blazed diffraction gratings to provide high efficiency for visible (blazed at 350 nm) and NIR (blazed at 1000 nm) wavelengths. The QTH lamp was chosen for its high light output in the near infrared. The monochromator slit width corresponds to a 10 nm band for use with the visible grating and a 20 nm band for use with the infrared grating. Longpass filters were placed at the monochromator input to eliminate unwanted contributions from higher diffraction orders. Using lines from a krypton gas discharge spectral calibration lamp it was determined that the wavelength selection of the monochromator is accurate to better than 1 nm.

The light from the monochromator passes through a filter wheel providing two neutral density (ND) filters with optical density 0.5 (32% transmission) and optical density 1 (10% transmission) and a fully open position. This allows for selectable attenuation. A fiber coupler at the filter wheel output connects to a liquid light guide which is followed by a glass rod providing a vacuum sealed feed-through into the dewar. A folding mirror inside the dewar reflects the light emerging from the glass rod into a 5 cm diameter PTFE integrating sphere (SphereOptics). The output port of the integrating sphere is projected onto the detector plane. A 64.4 cm long, cold black baffle tube encloses the light path between the integrating sphere and the detector in order to minimize reflections and to prevent contamination by stray light. Aeroglaze Z302, a glossy black absorptive polyurethane coating was applied over the illuminated sides of the knife edged circular aluminum baffles and Aeroglaze Z306 (flat, non-glossy) was used on the back sides of these baffles and all other surfaces. The temperature of the light projection system inside the dewar was not actively temperature controlled but equilibrates to a temperature of about 180 K after several hours.

Care was taken to ensure that the detector and the reference photodiodes are illuminated uniformly. The illumination uniformity was verified by scanning a photo-

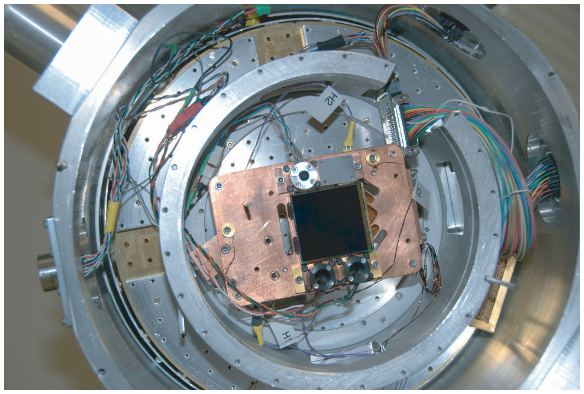


Figure 6.2: Picture showing the inside of the dewar with the H2RG-236 detector and the two reference diodes installed. A third photodiode is also shown though it is not used.

diode across the detector plane with light at 600 nm and 900 nm. The non-uniformity in the illumination was measured to be below ±2% and therefore does not constitute a significant source of uncertainty in the measurements presented here.

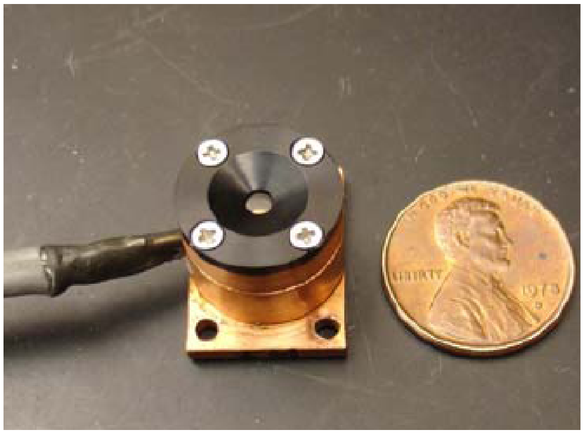


Figure 6.3: Picture of the photodiode package next to a penny for size comparison. The package was designed to operate at cold temperatures (100 K and 140 K). It features a copper housing, black anodized aluminum aperture, and padding that electrically isolates the package yet allows for good thermal conductivity.

6.1.2 Reference Photodiodes

Two calibrated reference photodiodes are used for the measurement of absolute quantum efficiency; a Silicon (Si) diode, Hamamatsu S1336-44BK, for measurements below 1000 nm and an indium gallium arsenide (InGaAs) diode, Hamamatsu G8373-03, for measurements at wavelengths above 800 nm[1]. The overlap in wavelength coverage between the two photodiodes allows for a cross check of their calibrations. The reference photodiode was placed as close as possible to the tested detector in order to minimize differences in photon flux and illumination uniformity. Given the roughly 4 cm × 4 cm size of a H2RG detector, co-locating the photodiodes next to the detector in a dewar requires that the photodiodes be in a small package and operated at focal plane tem-

---

[1] Throughout this article we will indicate whether measurements were referenced to the Si diode ('Si') or the InGaAs diode ('InGaAs').

perature along with the detector. The photodiodes are 3 mm in diameter and were mounted onto a copper pedestal to provide good thermal connection (Figure 6.3). A black anodized aluminum aperture of 1 mm diameter defines the sensitive area. The NIST calibrated photodiodes from which all calibration curves were transferred have a calibrated spectral irradiance response error of 1.5% to 1.95% between 700 nm and 1700 nm. Calibration transfer adds an additional uncertainty of about 1%.

Photodiode currents are monitored by a Keithley 2502 dual channel picoammeter and read out through a GPIB interface by the data acquisition computer. For stable performance, the picoammeter was turned on at least one hour prior to every series of measurements. Photodiode currents vary from approximately 0.1 pA at the blue end of the spectrum to approximately 14 pA in the NIR. An accurate measurement requires multiple samples. This is achieved by operating the picoammeter in a multi-sampling mode where the instrument averages over 5 independent current samplings. In addition, the computer triggers 10 such acquisitions and from those the average current and its statistical uncertainty was computed.

6.1.2.1 Absolute Calibration Transfer

To study temperature effects, QE measurements were performed at two temperatures, 100 K and 140 K. The temperature of 140 K is a typical operation temperature for the H2RG 1.7 μm cut-off detector and reduces dark current to insignificant levels in the majority of devices tested. The 100 K temperature point was chosen because previous measurements had shown that reciprocity failure at this temperature is significantly suppressed (Biesiadzinski et al. 2011a). Conveniently, the same dewar and illumination system used for the QE measurements could be used for calibration transfer. Both reference diodes were temperature calibrated by performing calibration transfer from two previously calibrated photodiodes. Those *standard* photodiodes had been calibrated to 140 K at Indiana University (Mostek 2007; Schubnell et al. 2008) using room-temperature NIST calibrated diodes.

In order to transfer the calibration to the reference photodiodes at 100 K and 140 K, each of the two standard diodes was mounted next to and thermally isolated from the corresponding reference diode at the detector plane inside the test dewar and independently temperature controlled. From the measured photo-currents at each monochromator wavelength a transfer function was calculated. Calibration transfer adds a negligible amount of uncertainty to the InGaAs diode calibration but it contributes substantially to the Si diode calibration uncertainty due to low light levels. Figure 6.4 shows the photodiode calibration curves (top panel) and corresponding

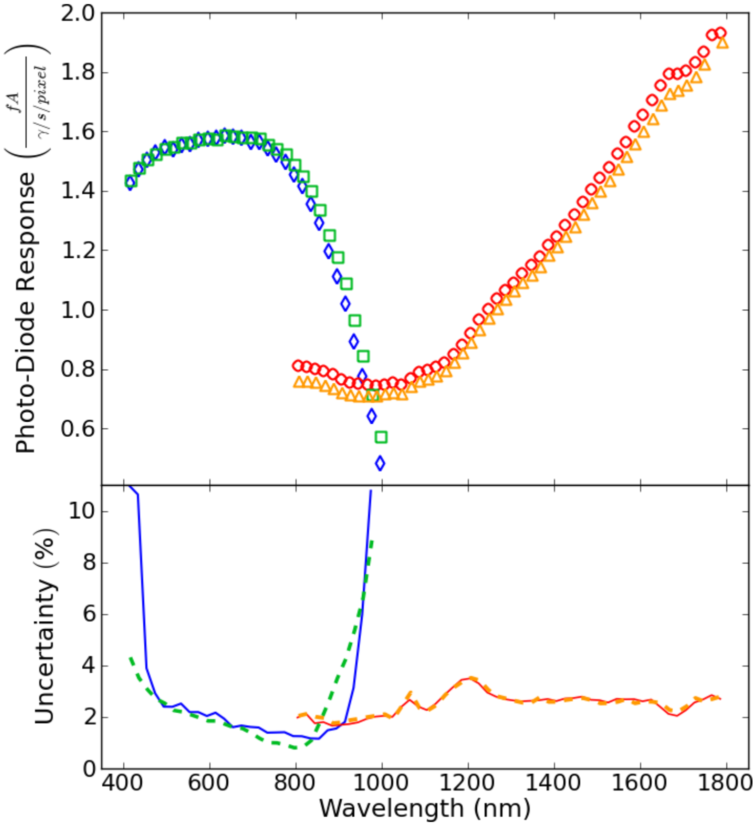


Figure 6.4: Top panel: Calibration curves for the reference photodiodes. Blue diamonds (Si) and red circles (InGaAs) indicate reference photodiode response at 100 K and green squares (Si) and orange triangles (InGaAs) show photodiode response at 140 K. Bottom panel: The systematic uncertainty in the absolute calibration expressed as a fraction of the measured photodiode response (color scheme as above.)

systematic uncertainties (bottom panel).

Calibration transfer for wavelengths below 450 nm could not be performed directly due to the low intensity of the QTH lamp at those wavelengths. Instead, calibration values were obtained by calculating the ratio of the 100 K calibration to the standard 140 K calibration between 450 nm and 700 nm and extrapolating it to wavelengths below 450 nm. In order to account for deviations from this extrapolation an additional 10% systematic uncertainty in the Si photodiode calibration was estimated and added to the total uncertainty.

6.1.2.2 Linearity

Photodiodes can suffer from flux dependent non-linearity (Biesiadzinski et al. 2011b) and therefore the linearity of the reference diodes must be characterized and corrected if necessary. For this measurement the two reference diodes were mounted side by side at the same position as during the QE measurement (see Figure 6.2) and a third photodiode was placed inside the integrating sphere to monitor variations in the light intensity. This *monitoring photodiode* was stabilized at a temperature of 270 K where its response non-linearity was previously measured to be less than 0.1% per decade in illumination difference. For the monitoring diode an absolute calibration is not required since flux dependent non-linearity is a relative effect. A correction factor was determined by comparing the photo-currents recorded by the two reference diodes with the photo-current measured by the monitoring diode at every measured wavelength.

Data was taken at wavelengths between $\lambda_c = 400\,\mathrm{nm}$ and $\lambda_c = 1800\,\mathrm{nm}$ and at three intensity levels, unattenuated, attenuated by a ND 0.5 filter and attenuated by a ND 1 filter. λ_c refers to the central wavelength of the 10 nm wide in visible and 20 nm wide in NIR monochromator slit width. This intensity range matches the range of illuminations used for the QE measurement. At each wavelength, the ratios of the reference photodiode currents at different attenuation levels were computed relative to the unattenuated illumination. The ratios were then scaled by the attenuation amount as determined by the responses of the monitoring diode. A value of unity represents a linear response in illumination intensity while any deviation from unity measures the diode illumination non-linearity per decade of change in illumination. The calibration is shown in Figure 6.5. The InGaAs reference diode does not show significant non-linearity over its sensitivity range. The Si diode however exhibits highly non-linear behavior towards the red end of its sensitivity range (above $\approx$ 800 nm).

The non-linearity was modeled to allow interpolation and estimation of uncertainties. A cubic polynomial was the lowest order polynomial that described the data well. Calibration uncertainties were scaled to account for deviations from the cubic parametrization near the diode cut-off and expanded below 600 nm where the data had little constraining power. The modeled non-linearities are shown in Figure 6.5 as black curves along with one standard deviation uncertainty bands. Measured photodiode currents recorded during the QE measurement were corrected for non-linearity according to this function. This improved the agreement between the device QE relative to the two photodiodes in the wavelength region where they overlap.

6.1.3 Electronics and Data Acquisition

For detector read-out and control, a commercial data acquisition system from Astronomical Research Cameras was used. In this system 32 channels of parallel read-out are available from four 8 channel infrared video processor boards combined with clock driver boards and 250 MHz timing and PCI cards. A detailed description of the read-out electronics can be found in Leach & Low (2000b). A Python script was

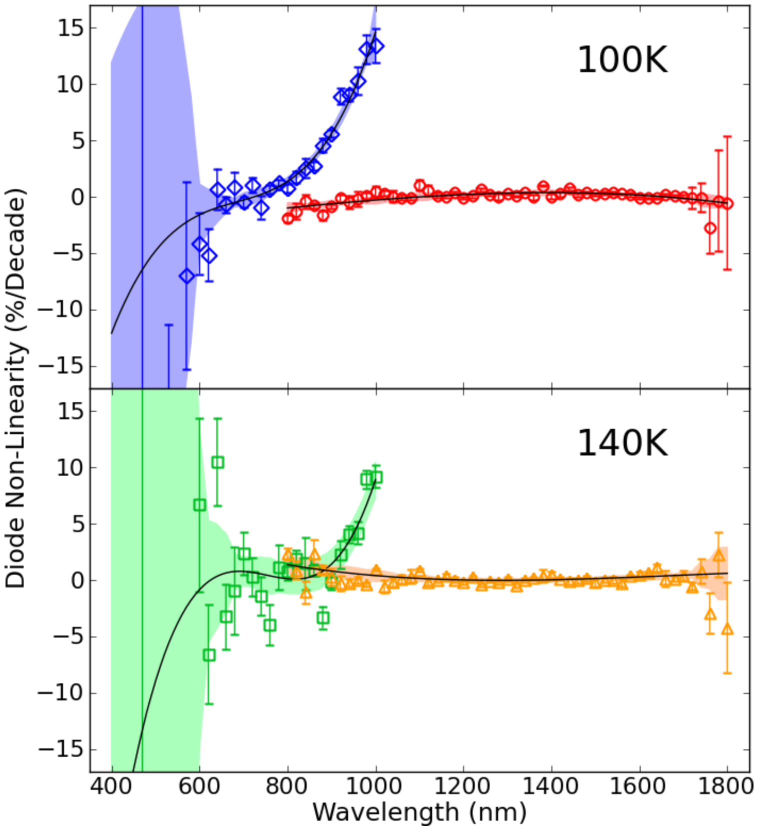


Figure 6.5: Si and InGaAs reference photodiodes non-linearity per decade of change in illumination intensity. Top panel: Blue diamonds (Si) and red circles (InGaAs) show reference photodiode non-linearity at 100 K. Bottom panel: Green squares (Si) and orange triangles (InGaAs) show the reference photodiode non-linearity at 140 K. The black solid curves show the non-linearity model. The shaded colored regions represent the 1σ uncertainty of this model.

developed to automate the data acquisition process. This reduced operator errors and eased operation and control.

6.2 Measurements

A near-infrared detector (H2RG-236) with large reciprocity failure was selected for the measurements to emphasize its effect on laboratory quantum efficiency measurements. The characterization of the reciprocity failure of this device was reported in Biesiadzinski et al. (2011a). It was measured to be about 10% per decade of illumination change and to saturate at very high and low flux.

The H2RG 236 detector is a 1.7 μm cut-off, substrate removed HgCdTe array that was produced during the 6^{th} manufacturing run for the SNAP project in 2008 by TIS. That production run addressed the issue of capacitive coupling by hybridizing the detector to a slightly modified version of their standard HAWAII-2 read-out multiplexer. This detector has an anti reflective coating and is mounted onto a silicon carbide (SiC) pedestal. Manufacturing and processing of the detector largely followed the recipe that produced the excellent flight detector for the WFC3 team (Baggett et al. 2008). Dark current and read noise performance for this detector are good, with a correlated double sampling (CDS) read noise of 25.7 electrons (e^-) and a dark current of 0.03 e^-/s. While a slight improvement of the inter-pixel capacitive coupling was achieved, the most notable performance difference to previously produced devices was an increase in persistence by almost a factor of 10 (Roger Smith, personal communication, 2008).

6.2.1 Wavelength Spacing

Measurements were taken over the visible and near-infrared wavelength range within a 20 nm wide window centered at wavelengths $\lambda_c = 400$ nm to $\lambda_c = 1800$ nm. A sampling step size of 20 nm was chosen to match the monochromator slit width of 20 nm needed to obtain intensities sufficiently large to complete a single measurement at all wavelengths within several hours. Because of the low light output of the QTH lamp in the visible part of the spectrum, the signal to noise (S/N) ratio in this range is low. In order to improve S/N, measurements were grouped in the visible part of the spectrum (390−550 nm, 550−690 nm, 690−790 nm, 790−850 nm and 850−910 nm).

6.2.2 Operating Temperatures

Measurements were performed at 100 K and 140 K. For a 1.7 μm cut-off device an operating temperature of 140 K is typically sufficiently low to reduce the dark current to insignificant levels. However, for very long exposure times lower operational temperatures may be required. It is therefore of interest to understand how QE will be affected by device temperature. More importantly, it was observed that reciprocity failure is a strong function of temperature (Biesiadzinski et al. 2011a) and measurements on several H2RG devices showed that reciprocity failure can be largely eliminated at operating temperatures of 100 K.

6.2.3 Illumination Levels

We investigated reciprocity failure for H2RG-236 by varying the illumination intensity. The output spectrum of the lamp, and to a lesser degree, monochromator efficiency and atmospheric absorption limit the ability to maintain a constant level of illumination at all wavelengths. Therefore three distinctly different illumination levels with constant illumination ratios were produced using ND filters. We compare the device response at the following illumination levels: *high* illumination (no attenuation), *medium* illumination (attenuation by a factor of 3.2) and *low* illumination (attenuation by a factor of 10).

6.2.4 Integrated Signal Levels

To investigate the impact of the integrated signal non-linearity measurements were performed at *deep* and *shallow* absolute integrated signal levels with total integration levels of approximately 28,000 e^- and 7,000 e^-, respectively. This corresponds to a well fill of roughly 28% and 7%, respectively. The measurement was performed at *medium* illumination (see §6.2.3) to prevent reciprocity failure effects. The *deep* integration can only be maintained at wavelengths above 800 nm due to the low light output of the QTH lamp at shorter wavelengths and therefore the analysis is limited to those wavelengths. All reciprocity failure measurements were performed at *deep* integration.

6.2.5 Conversion Gain Factor

The absolute quantum efficiency is obtained through a comparison of the accumulated charge at the NIR detector unit cell and the photon flux measured by the reference photodiodes. The detector readout however is sensitive to voltage change (in units of analog to digital units (ADU)) across the pixel capacitance induced by the collected charge. It is therefore necessary to convert the voltage recorded in ADU to electron count. We do so by using Poisson counting statistics of electrons collected on the device to determine the conversion gain factor (Mortara & Fowler 1981; Janesick et al. 1985). The variance in the recorded signal, in units of electron counts squared, then equals the mean signal in electron counts plus an offset due to other noise sources.

The matter is complicated in the presence of capacitive coupling between adjacent pixels (Moore et al. 2004; Finger et al. 2005; Moore 2006; Brown et al. 2006). Adjacent pixels share a small fraction of the difference in their charges. This has the effect of reducing the measured spatial variance and therefore artificially increasing the measured conversion gain factor. Moore (2006) gives a variance estimator that accounts for correlations between adjacent pixels and allows for the recovery of an unbiased conversion gain factor. Mean signal and variance then follow the relation

$$\sigma^2_{ADU} = \frac{\mu_{ADU}}{\kappa} + \eta \ , \tag{6.1}$$

where σ^2_{ADU} is the correlation-corrected variance in units of ADU, μ_{ADU} is the mean signal in units of ADU, κ is the conversion gain factor in units of e^-/ADU and η is extra noise variance from non-Poisson sources independent of flux or exposure time (e.g. read noise). Multiple measurements are taken where the signal is integrated to different mean total counts μ_{ADU}. The above relation is fit to the data to obtain an estimate of the conversion gain factor. The method suffers from systematic problems related to the nature of the measured variance. That is, the variance at any μ_{ADU} can be larger than the Poisson-expected variance due to additional sources of noise (e.g. detector response gradients and temporal noise). For our analysis, several different algorithms were applied to the data to estimate the conversion gain factor. Some focused on improving the overall variance measurement quality, at the cost of possible small negative biases in the conversion gain factor, while others explored minimizing variance contamination by non-Poisson processes leading to a possible small positive bias in the conversion gain factor. Averaging over the results obtained through those methods resulted in a conversion gain factor of $1.75 \pm 0.06\,e^-$/ADU at 100 K and $1.73 \pm 0.06\,e^-$/ADU at 140 K for this device. The uncertainty estimate captures the systematic errors in our analysis.

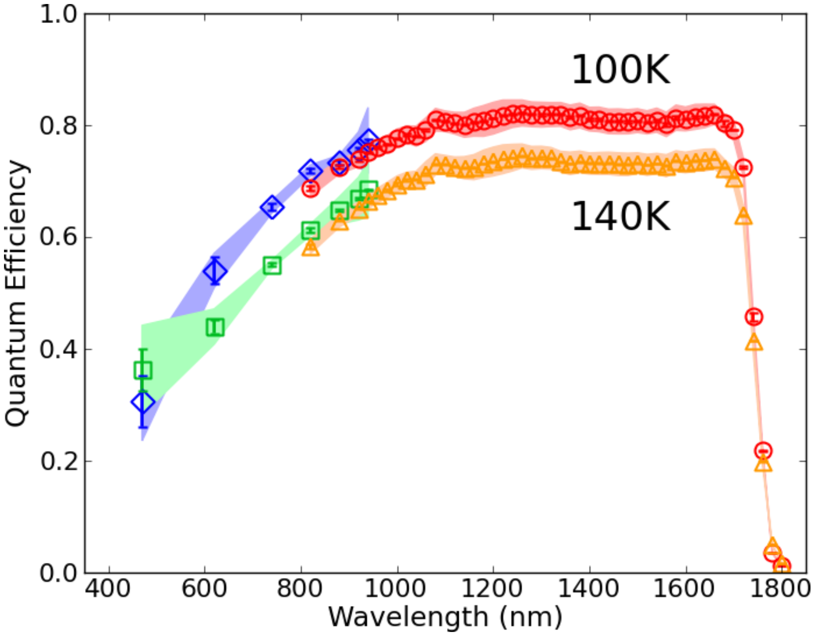


Figure 6.6: Quantum Efficiency as a function of wavelength at 100 K and 140 K. Measurements were performed at *high* illumination and *deep* integration. Statistical uncertainties ($1\,\sigma$) are shown as error bars and total uncertainties (statistical and systematic) are indicated by shaded regions. Blue diamonds (Si) and red circles (InGaAs) show the QE at 100 K and similarly, green squares (Si) and orange triangles (InGaAs) show measurements at 140 K.

It is interesting to note that while the conversion gain factors measured at the two temperatures are consistent, the correction factor for capacitive coupling was approximately 1.8 at 100 K, but only 1.2 at 140 K. This indicates an increase of inter-pixel capacitance at lower temperatures. This observation is consistent with measurements reported by Cheng (2009).

6.3 Results and Discussion

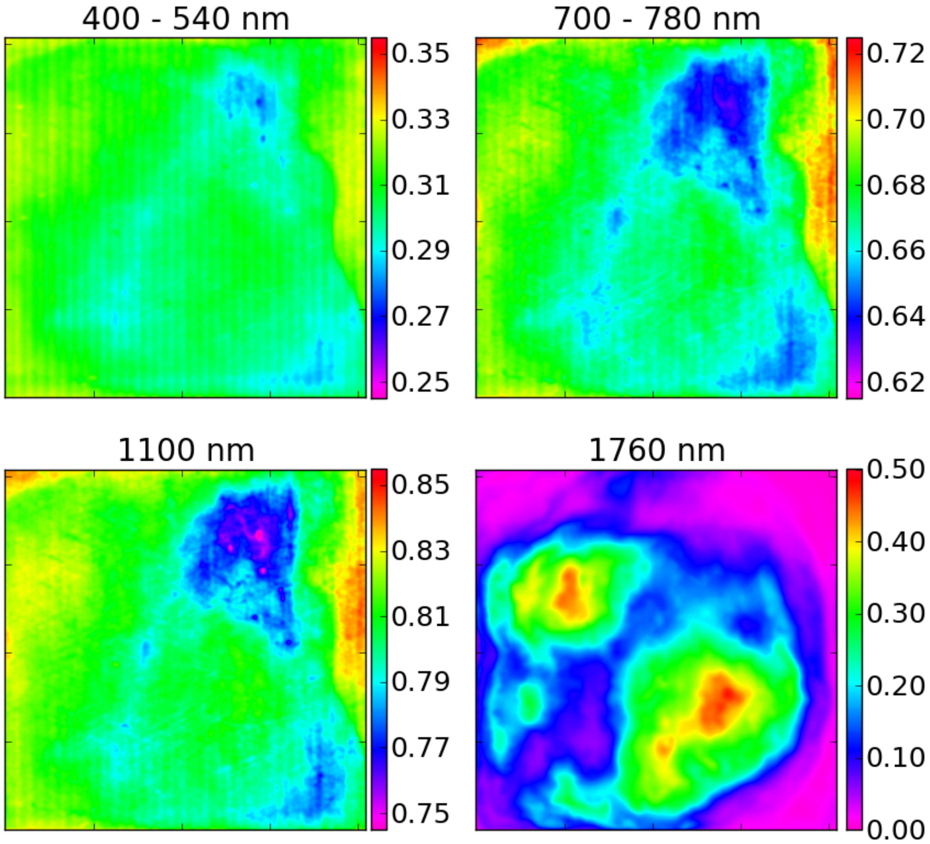


Figure 6.7: QE maps at 100 K for four wavelength regions. Combination of data from 400 nm to 540 nm (top left), data from 700 through 780 nm (top right) 1100 nm data (bottom left) and 1760 nm data (bottom right). Note that the QE maps were smoothed with a Gaussian kernel with full-width-half-max of 20 pixels in order to suppress shot noise and to emphasize large scale structure.

The detector averaged QE measured at 100 K and 140 K is shown in Figure 6.6. This data was taken at *deep* integration and *high* illumination. The statistical and systematic uncertainties are large in the blue due to the low light of the QTH light source at those wavelength. QE values in the region of overlap (800 nm to 950 nm) show a small systematic mismatch, well within the systematic uncertainties due to the original photodiode calibration. QE maps showing the spatial variation of QE across the detector at 470 nm, 740 nm, 1100 nm and 1760 nm are shown in Figure 6.7. Significant structure becomes apparent at 1760 nm, near the device cut-off wavelength.

6.3.1 Uncertainty Budget

Great care was taken to account for known sources of statistical and systematic uncertainties. As an example the systematic uncertainty decomposition for the *deep* integration measurements at 100 K is shown in Figure 6.8 along with the total statistical uncertainty. Uncertainties at 140 K are in general very similar though the contribution due to photodiode flux non-linearity is somewhat larger because of the larger non-linearity measurement uncertainties (see Figure 6.5).

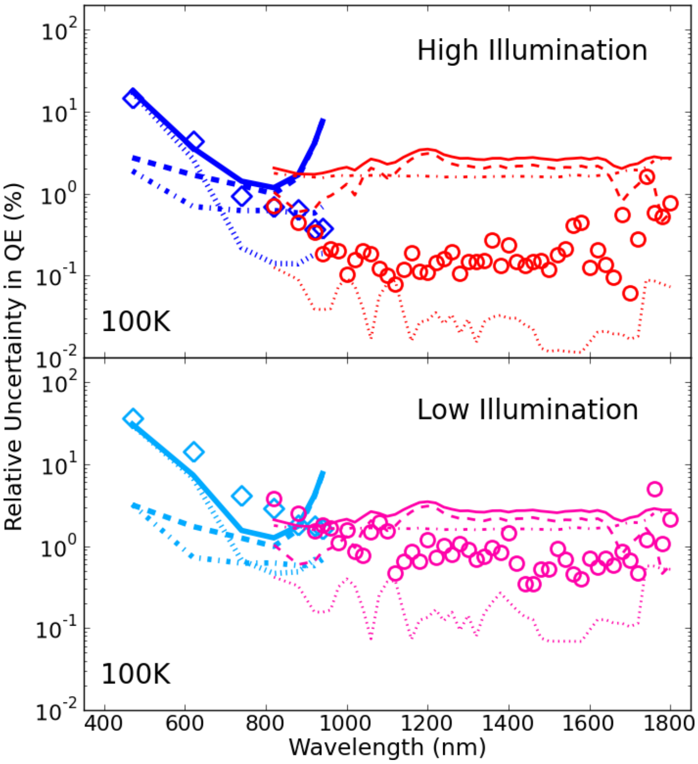


Figure 6.8: Relative uncertainty in QE at 100 K for *deep* integration measurements. Top panel: *high* illumination data; blue diamonds (Si) and red circles (InGaAs) show the statistical uncertainties; solid blue (Si) and red (InGaAs) curves show the total systematic uncertainties. The systematic uncertainties are further resolved into contributions from original photodiode absolute calibration errors (dashed), an estimated 10 nm calibration wavelength uncertainty (dot-dashed) and photodiode flux non-linearity (dotted). Bottom panel: *low* illumination data; the above color scheme is replaced using cyan (Si) and magenta (InGaAs).

Although significantly more time was spent recording data at *low* illumination, the statistical uncertainties for those measurements are significantly higher than for those at the *high* illumination and are dominated by the photodiode current noise. The second largest contribution is the electrical bias drift. Other identified sources of statistical error are the 1 nm wavelength jitter error and the detector shot noise, both of which are usually smaller than 0.05%.

The systematic uncertainties are composed of the original absolute calibration uncertainty, inferred calibration wavelength error of 10 nm and the photodiode non-linearity uncertainty which only contributes significantly below 700 nm. The non-linearity contribution is larger at *low* illumination which deviates further from the nominal calibration transfer current.

Additionally, a constant 3.4% systematic uncertainty at all wavelengths, integration and flux levels due to the conversion gain factor estimate has to be taken into consideration (see Sec. 6.2.5). This contribution is not included in Figs. 6.6 and 6.8.

6.3.2 Charge integration non-linearity

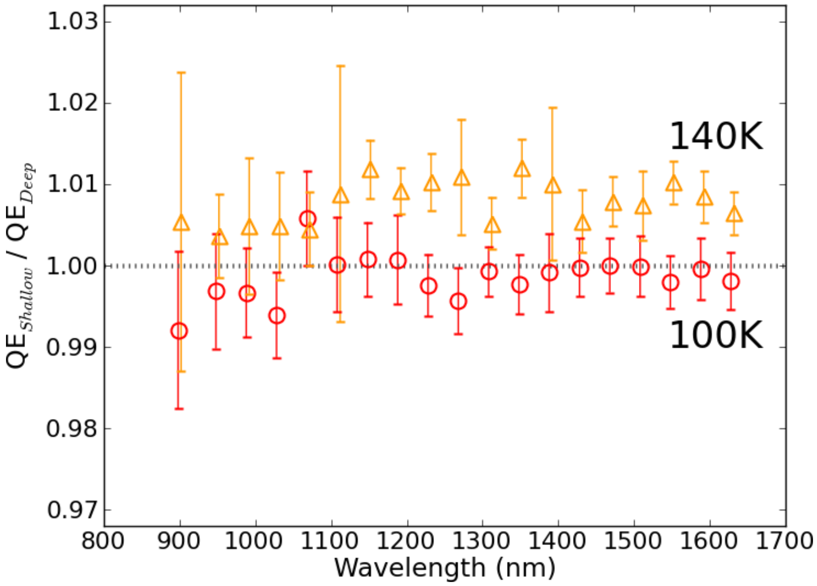


Figure 6.9: Ratio of the measured QEs at *shallow* and *deep* integration levels (100 K –red circles and 140 K – orange triangles). The dotted gray line marks the absence of charge integration non-linearity. Both measurements were performed at *medium* illumination.

Medium illumination data at *shallow* integration and *deep* integration were compared to study effects of integrated signal non-linearity. The analysis was limited to data obtained with the InGaAs photodiode in the wavelength range 850 nm – 1650 nm, where the total integrated signal is constant to within ± 4%. The results of these measurements are shown in Figure 6.9.

The cause for the integrated signal non-linearity is the dependence of the pixel node diode junction capacitance on the voltage applied to that junction as photoelectrons are collected. An increase in the voltage reduces the capacitance, causing the detector to record an apparently smaller flux. Therefore the *shallow* integration QE is expected to be higher than the *deep* integration QE since the *shallow* integration level corresponds to a lower junction voltage.

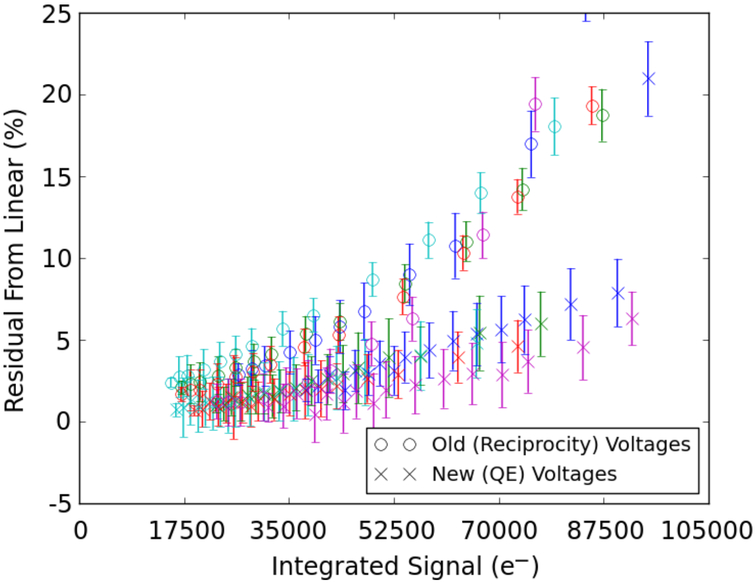


Figure 6.10: Integrated signal non-linearity of old (reciprocity failure) and new (QE) bias voltages. The different colors represent five different regions in the device selected randomly to check for spatial variation. Measurement performed at 1060 nm.

At 100 K the measured *deep* integration level QE was 0.1% higher than the *shallow* integration QE, indicating that the charge integration non-linearity is negligible. It is consistent with zero at the 2.5 σ level. At 140 K the the *shallow* integration QE was measured to be 0.8% larger than the *deep* integration QE. This level of charge integration non-linearity is smaller than the value of 2.5% expected from Biesiadzinski et al. (2011b). This difference originates in a different set of detector bias voltages used for the two measurements as shown in Figure 6.10. The QE was measured with an applied 0.5 V bias voltage across the pixel capacitance while the previous reciprocity failure measurement was made with the bias voltage set to 0.25 V. The larger voltage is likely to make the pixel capacitance less sensitive to the total integrated charge.

6.3.3 Impact of Reciprocity Failure

The impact of reciprocity failure on QE as a function of wavelength was measured by varying the illumination intensity over an order of magnitude. For this QE was sampled at the three illumination levels, *low*, *medium* and *high*. In Figure 6.11 the ratios of *high* to *medium* illumination QE and *low* to *medium* illumination QE at 100 K (top panel) and at 140 K (bottom panel) are shown. Each ratio spans approximately half a decade in illumination intensity. At 140 K the QE increases with increasing illumina-

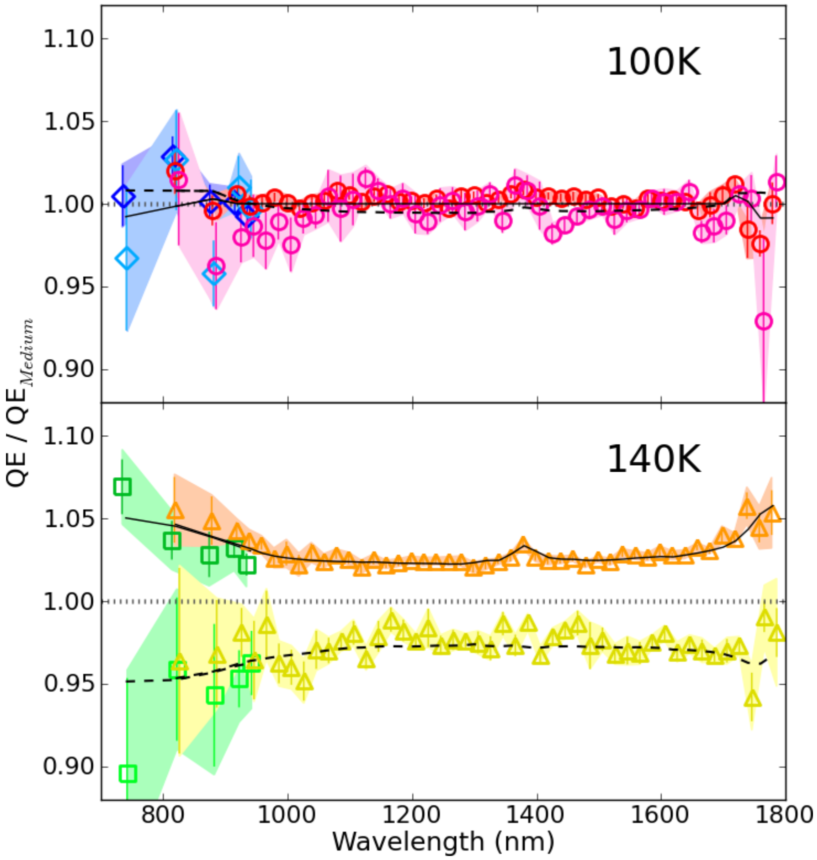


Figure 6.11: Ratios of QE measured at various illumination levels to QE measured at *medium* illumination. Top panel: T=100 K; blue diamonds (Si) and red circles (InGaAs) show the comparison of *high* illumination to *medium* illumination. Cyan diamonds (Si) and magenta circles (InGaAs) show comparison of *low* illumination to *medium* illumination. Bottom panel: T=140 K; green squares (Si) and orange triangles (InGaAs) show comparison of *high* illumination to *medium* illumination and light green squares (Si) and yellow triangles (InGaAs) show comparison of *low* illumination to *medium* illumination. For both panels statistical uncertainties of 1 σ are shown as error bars and total uncertainties (statistical and systematic) are indicated as shaded bands. The dotted black line marks the absence of reciprocity failure. Solid and dashed black lines show the ratios expected from the model fits at each temperature shown in Figure 6.13 for *high* illumination and *low* illumination ratios, respectively.

tion intensity as indicated by the positive ratio in the bottom panel of Figure 6.11 due to *high* illumination and the negative ratio in that panel due to *low* illumination. The effect is significantly smaller at 100 K as expected from Biesiadzinski et al. (2011a). Reciprocity failure for this device was measured to be slightly larger than 10% per decade in illumination (normalized at 100 e^-/s) hence one would expect the ratios in Figure 6.11 to be approximately ±5% at 140 K. However, at most wavelengths a value of about ±2.5% is measured and only at wavelengths between 1700 nm and 1800 nm and below about 900 nm does the measured value agree with previous measurements. Wavelength dependence was not expected from Biesiadzinski et al. (2011a).

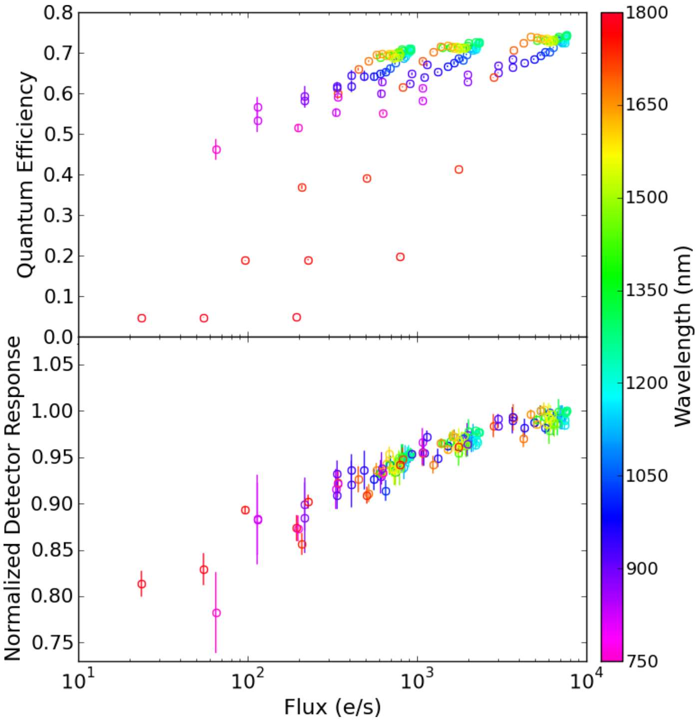


Figure 6.12: Application of the QE normalization procedure to recover reciprocity failure at 140 K. Top panel: QE plotted vs the flux seen in the device color-coded by the wavelength of the measurement. There are three fluxes per wavelength at high, medium and low illumination intensities. Bottom panel: QE from the top panel normalized to the highest flux measured. Error bars here account for uncertainties due to the normalization procedure.

The apparent inconsistency of this measurement with previous measurements of reciprocity failure for this detector prompted further studies. For the measurement of reciprocity in Biesiadzinski et al. (2011a) a dedicated illumination system allowed varying detector irradiance over 5 orders of magnitude. This is different from the QE set-up where control over irradiance is limited. However, the combination of wavelength dependent lamp intensity and the selectable ND filters provide a reasonably large sample of flux levels (measured in e^-/s/pixel) to perform a coarse measurement of detector response vs flux, i.e. reciprocity failure. To remove wavelength dependency from the QE measurement data, the following method was used: The *high* illumination QE data point at the wavelength corresponding to the highest flux was chosen as reference point and the responses at *medium* illumination and *low* illumination at this wavelength were normalized to the reference point. This provides data points of detector response at three different flux levels. In successive steps, the algorithm selects QE values (*high*, *medium*, and *low*) obtained at different wavelengths but with a flux value closely matching the flux of the previously selected data and the detector response again is normalized. This iterative method results in a set of normalized detector response measurements at different flux values. It is visually represented in Figure 6.12 for 140 K data where the top panel shows the QE values vs. the flux with wavelength encoded by the color-scale. The bottom panel shows the results of our procedure. In order to reduce the statistical uncertainty, multiple data points were combined. The results are shown in Figure 6.13. At 100 K, as expected, non-linearity due to reciprocity failure is not observed. At 140 K however, strong reciprocity failure is evident in the data. The monotonically increasing response as a function of flux further suggests that wavelength dependence is in fact not present in the data.

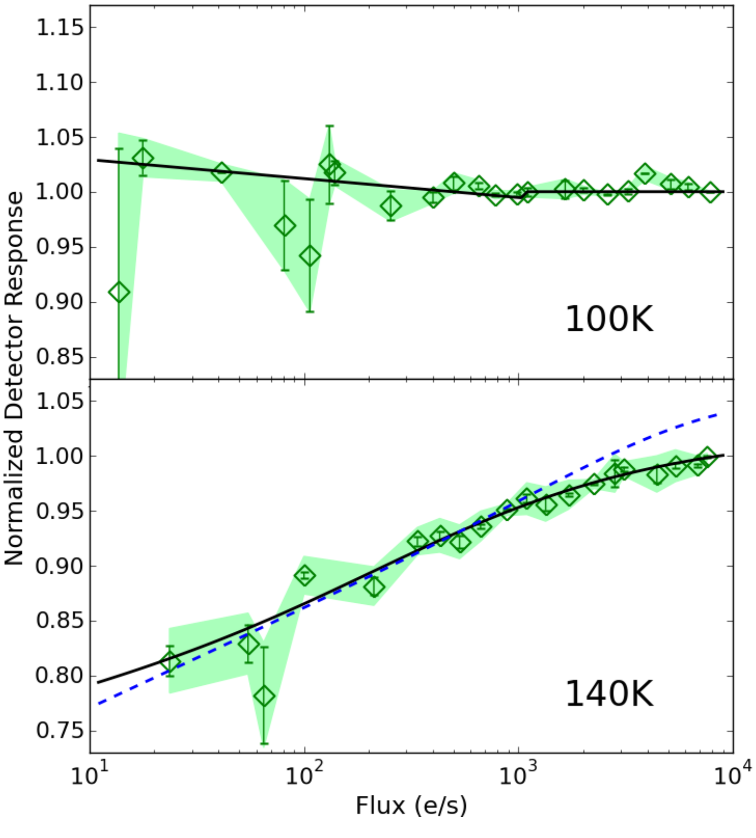


Figure 6.13: Relative detector response vs. intensity (in e^-/s/pixel) derived from QE measurements at 100 K (top panel) and 140 K (bottom panel). Statistical uncertainties of $1\,\sigma$ are shown as error bars and total uncertainties are indicated as shaded bands. Best fit curves (black solid lines) are shown for each temperature. In the bottom panel a blue dashed line indicates the reciprocity failure results from Biesiadzinski et al. (2011a).

Best fits to the data are shown in Figure 6.13 as black curves. The reciprocity failure result from Biesiadzinski et al. (2011a) is indicated as a dashed blue curve (bottom panel) for comparison. The best fit model results are also shown in Figure 6.11 where they reproduce the apparent wavelength dependence of QE ratios at 140 K

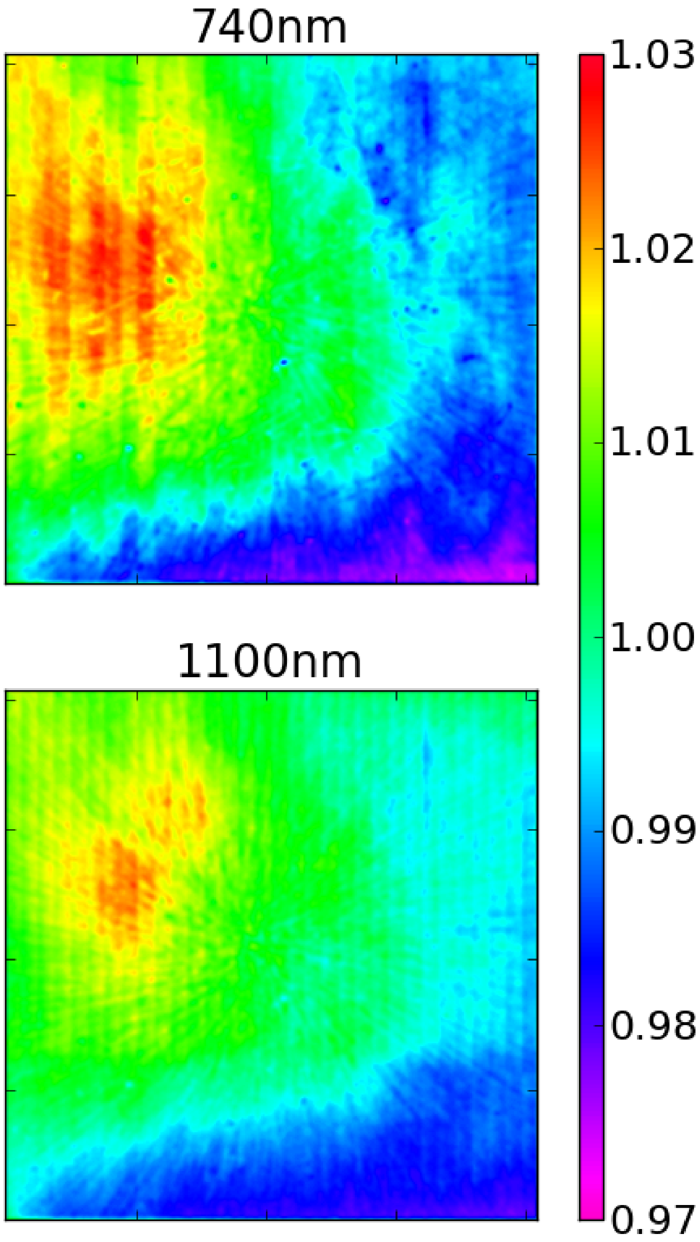


Figure 6.14: QE illumination dependence at *deep* integration, T=140 K and λ_c=740 nm (top panel) and λ_c=1100 nm (bottom panel). Maps show the ratio of measured QE for *High* illumination and *low* illumination, normalized to unity median. Reciprocity failure structure is clearly visible. Note that maps were smoothed with a Gaussian kernel (FWHM 20 pixels) in order to suppress shot noise and highlight large scale structure.

due to the wavelength dependence of illumination intensity even though reciprocity failure itself has no wavelength dependence. Comparing the solid black and dashed blue lines in the bottom panel of Figure 6.13 it appears that reciprocity failure in the QE data saturates sooner or at least becomes smaller than previously measured at fluxes larger than 1000 e^{-}/s but remains largely unchanged at lower fluxes. This hints at the possibility that an increase in the bias voltage (as discussed in §6.3.2) may be responsible for this earlier saturation in reciprocity failure if it can attenuate current leakage that could be responsible for reciprocity failure. If this is shown to be true by future measurements, it could provide a way to better control or even eliminate reciprocity failure. Other issues however, like increased detector dark current due to increased voltages, must be considered.

In addition, Figure 6.14 shows using the QE data that the expected reciprocity failure pattern is present. This figure shows the ratios of the *high* illumination to *low* illumination QE maps at 740 nm and 1100 nm. These measurements were obtained at 140 K and *deep* integration. These maps are qualitatively similar to the previously obtained reciprocity failure map (Biesiadzinski et al. 2011a). Similar ratios at 100 K have no significant structure and are not shown.

6.3.4 Temperature dependence

A significant drop in the average detector QE was observed when increasing the detector temperature from 100 K to 140 K (see Figure 6.6). This change in QE is not uniform across the detector. Figure 6.15 shows the spacial structure in the QE for measurements at 1100 nm and at 100 K and 140 K. We investigated whether the difference in QE between the two temperatures shows correlation with reciprocity failure and thus could be explained by the temperature dependence of reciprocity failure.

In Figure 6.16 the ratio of QE at 100 K to 140 K is shown for three different wavelengths. The patterns are consistent with those observed in reciprocity failure maps (see Figure 6.14) with the exception of an approximately 500 pixel × 500 pixel region in the top right of the image shown in the top panel of Figure 6.15. This area appears as a strong signal deficit and indicates a location where the detector at 100 K is quite insensitive. The reciprocity failure pattern itself appears to be larger in the bluer parts of the spectrum. This is consistent with the fact that the slope of the non-linearity is greater at lower fluxes. Results from an additional test designed to determine if reciprocity failure can indeed explain the temperature dependence of

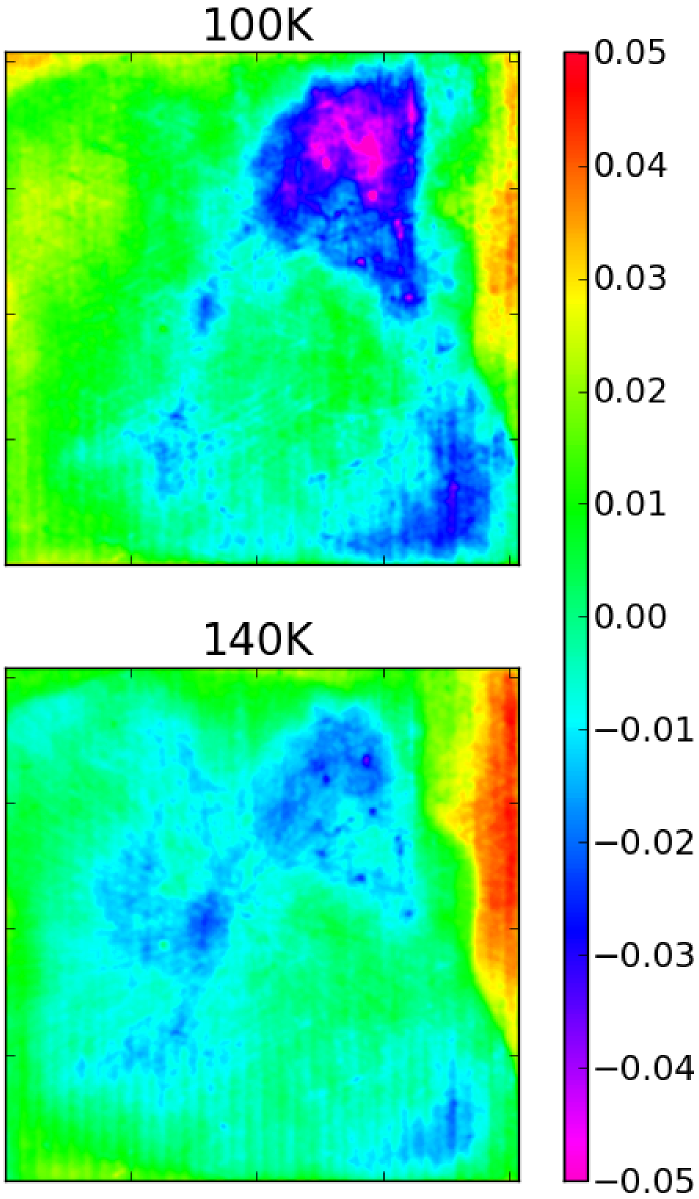
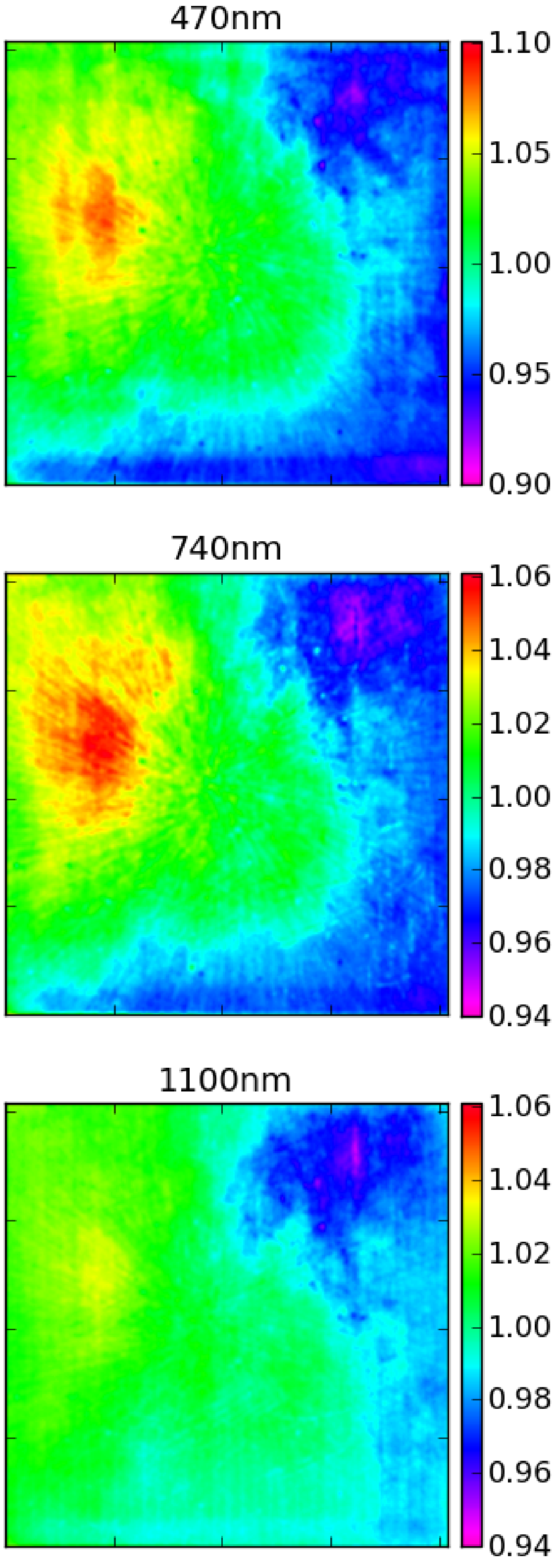


Figure 6.15: QE residual from mean maps at λ_c=1100 nm and T=100 K (top panel) and T=140 K (bottom panel). Measurements were performed with *high* illumination and *deep* integration. Spatial response variations are noticeable. Maps were smoothed with a Gaussian kernel (FWHM 20 pixels) to suppress shot noise and emphasize large scale variations.


470nm

740nm

1100nm

Figure 6.16: QE temperature dependence at λ_c=470 nm (top panel), 740 nm (middle panel) and 1100 nm (bottom panel). Maps show the ratio of QE at 100 K to QE at 140 K, normalized to unity median. Measurements were performed with *high* illumination and *deep* integration. Note that maps were smoothed with a Gaussian kernel (FWHM of 20 pixels) in order to suppress shot noise and highlight large scale structure.

absolute QE is presented in Figure 6.17. In this figure the QE ratio of *high* illumination

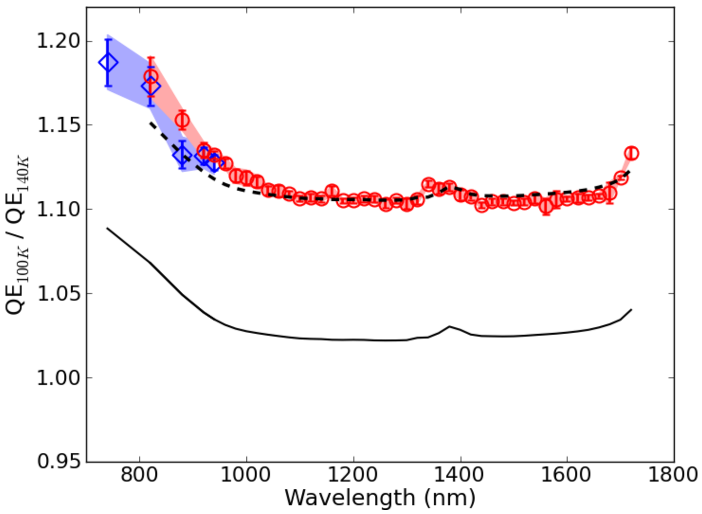


Figure 6.17: Ratio of *high* illumination, 100 K QE to *high* illumination, 140 K QE using Si photodiode shown as blue diamonds and InGaAs photodiode as red circles. Uncertainties of 1 σ are shown as error bars (statistical) and shaded regions (total). A prediction of the ratio using the reciprocity failure model from §6.3.3 is shown as a black line. It is also shown as a black dashed line after being manually raised to overlap the measured ratio. This illustrates that it explains most, though not all, of the apparent wavelength dependence. Data above 1700 nm is not shown since the device begins to cut off there and the cut-off itself experiences temperature dependence.

at 100 K to *high* illumination at 140 K is shown. The reciprocity failure model derived from QE measurements in Figure 6.13 was used to predict this ratio. This prediction is shown as a solid black line in Figure 6.17. There is clearly an overall offset. The model assumes that the reciprocity failure saturates at about 5,000 e^-/s and since the flux at 140 K in the NIR part of the spectrum is already that high, the model does not predict the observed overall increase. If instead, reciprocity failure does not saturate but continues, albeit at a reduced slope, towards some higher saturation value, it may explain this temperature ratio. It should be noted that reciprocity failure does explain the apparent increase below 1000 nm and the small bump around 1380 nm where the flux is lower due to water vapor absorption lines. The device response cut-off changes between the two temperatures (see §6.3.5), hence the ratio cannot be attributed to reciprocity failure above 1700 nm. Considering the QE map changes and this temperature ratio it appears that while reciprocity failure contributes to the overall temperature dependence of the device QE, it may not be the only source of the difference. Additional studies with a much larger range of illumination intensity would be required to confirm this.

6.3.5 NIR Cut-Off Shapes

The detector response for individual pixels at the NIR cut-off wavelength, defined as the wavelength where the QE is 50% of its prior steady state value, was investigated in detail during the QE study presented. Typically, filter transmission bands define the range of wavelengths to which a detector is exposed to. During previous QE measurements on several 1.7 μm cut-off HgCdTe detectors it was observed that near the cut-off QE does not decrease uniformly across the detector. For all detectors that were studied, pronounced large-scale variation appeared in the QE maps within the roll-off region. The cause for this was speculated to lie in inconsistencies in the stoichiometry of the material (Brown 2007). Our investigation focused on the question of how much the cut-off characteristics varies from pixel to pixel. This is of interest in applications where it is desirable to maximize the usable waveband of a detector without introducing non-uniformity in the detector response.

For the study, a pixel mask was produced and all dead or hot pixels as well as pixels in the vicinity of cosmetic irregularities (scratches etc.) were excluded in the analysis. This resulted in 8% of all pixels being masked. Next, an error function was fitted to the measured QE values between 1620 nm and 1800 nm for 10,000 randomly selected pixels. Note that the error function was chosen because it qualitatively resembled the QE transitions and not because of an underlying physical reason. It was meant to systematically center the pixel cut-off transitions and provide an estimate of the transition width. The initial fitted transition center wavelength was used to select only the QE data at the cut-off (5 points in wavelength spanning approximately 100 nm) for each pixel which was then re-fit to avoid biases due to the fitting procedure involved; for example, the pixels that “turn off” last have fewer wavelength data points available after the transition so the fitter may not constrain the model in the same way as for pixels that turn off earlier if the fit region is not limited.

It was determined that for the detector studied the majority of the cut-off transition centers are located between 1700 nm and 1760 nm and the map of these centers closely resembles the QE map observed at 1760 nm (see Figure 6.7). This is not surprising since at that wavelength most pixels are clearly “turning off” while others have already passed the cut-off. Interestingly, it was observed that the roll-off shapes fall in two populations (left panel of Figure 6.18): *narrow* and *broad*. By selecting pixels in the tails of the distribution and averaging over a large number of pixels the two representative shapes emerge in the right panel of 6.18. For pixels with *narrow* roll-off the QE drops quickly and smoothly from the peak value to zero while for pixels with *broad* roll-off the rate in QE drop increases at the drop-off midpoint and

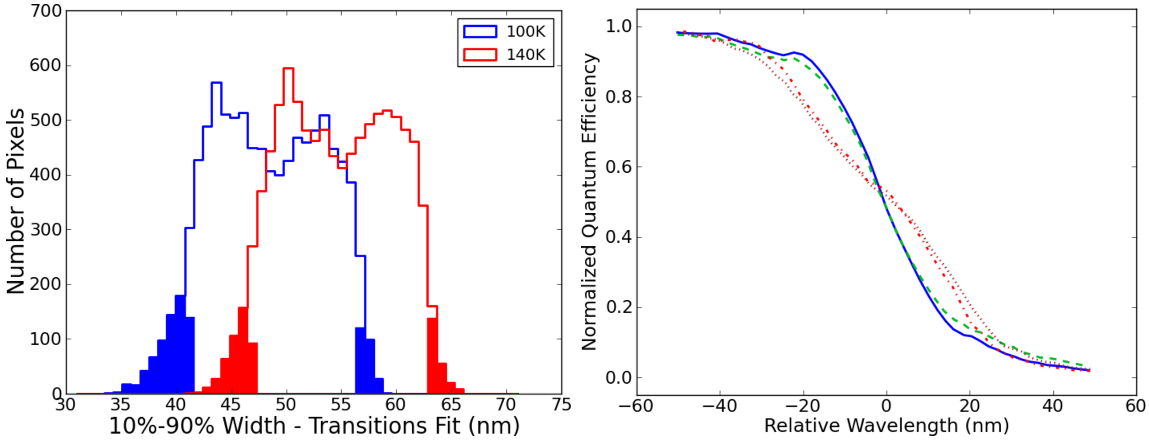


Figure 6.18: NIR cut-off populations and shapes. Left panel: the overlapping populations of *narrow* and *broad* pixel transitions at 100 K (blue) and at 140 K (red). Only pixels from the solid colored regions were used to produce the representative transition shapes shown in the right panel. Right panel: shown are averaged *narrow* transitions at 100 K (blue) and 140 K (cyan) and averaged *broad* transitions at 100 K (red) and 140 K (orange). All transitions are normalized to an amplitude of unity.

the transitions are, on average, broader.

Temperature dependence affects the overall QE over the entire wavelength range and in addition, at 140K the transitions (defined as the distance between 90% and 10% QE levels) tend to be 6 nm wider than at 100 K as can be seen in the left panel of Figure 6.18. The transition centers are shifted, on average, 1 nm towards the red. Overall, the temperature dependence of the cut-off transition does not appear to be significant.

6.4 Conclusions

A detailed study of quantum efficiency with the H2RG-236 1.7μm cut-off near-infrared HgCdTe detector was performed. As part of this investigation a precision QE measuring setup was constructed. Two reference photodiodes, calibrated at 100 K and 140 K were read out simultaneously with the detector. Their illumination level non-linearity was characterized. The Si photodiode showed significant flux non-linearity near its red cut-off. This could lead to a large errors in QE measurement at these wavelengths if not corrected for. In addition, uncertainties in the measurement of the conversion gain factor were characterized. It was found that capacitive coupling between pixels increases with decreasing temperature.

The study confirms previous measurements of reciprocity failure in the tested detector. It was also confirmed that reciprocity failure can be significantly suppressed by lowering the detector operating temperature from 140 K to 100 K. However, the measured reciprocity failure appears to saturate at a lower illumination intensity than previously measured. It is possible that the higher detector bias voltage used for the current measurements is responsible for this. It was found that the bias voltage change did reduce the integrated signal non-linearity. The QE for device H2RG-236 is significantly higher at 100 K than at 140 K. This difference can in part be explained by the measured reciprocity failure. However, general QE temperature dependence cannot be excluded. Finally, the pixel NIR cut-offs appear to be drawn from two overlapping distributions in width with distinct shapes.

CHAPTER 7

Instrumentation Summary

We have entered an era of precision cosmological studies where we aim to distinguish between multiple models of the universe and attempt to explain the mysterious dark energy. This can only be achieved if observational tools and techniques deliver data sets of unprecedented precision and accuracy. Semiconductor based detectors like CCDs have revolutionized the field of astronomy and their performance has been sufficiently well understood up until now. However, the era of precision astronomy and cosmology sets more stringent requirements on their performance. Current studies depend more and more on high redshift observations that require thicker CCDs as well as near-infrared (NIR) detector technology which is not as proven and understood as its optical counterparts.

Standard detector characterization includes the measurement of quantum efficiency (QE), dark current, read noise and integrated signal non-linearity. High QE improves the general observational performance by allowing more photons to be recorded by generating electrons. Dark current, i.e. thermally generated electrons in the bulk and surface detector material, and noise due to readout electronics constitute a portion of the noise from which the signal must be disentangled. Photon shot noise from sources, sky glow and zodiacal light constitutes the remainder of the noise. Dark current in the bulk detector material is known to be a function of the detector temperature and bias voltages while surface defects can contribute an additional, irreducible dark current floor. Read noise meanwhile is largely dependent on the readout electronics design and the readout speed. In addtion, both CCDs and NIR detectors are know to be non-linear with integrated signal, that is, the *effective* QE decreases as more charge is integrated in a given pixel. The effect is larger in NIR devices. CCDs can also suffer from issues related to the way they are read out such as dead columns, blooming (a charge spill over effect) and charge transfer errors (where charge can be lost or smeared during readout).

All of the above mentioned effects are know and generally considered when designing astronomical cameras. However, precision cosmology requires deeper understanding of detector response. Mercury-Cadmium-Telluride (HgCdTe) NIR devices in particular have been studied closely and have revealed a multitude of effects that may significantly alter their performance. This dissertation discusses flux dependent nonlinearity, reciprocity failure, (Chapter 5), detailed QE characterization under various conditions (Chapter 6) and subpixel structure (Appendix C). In addition, an effect called persistence affects these detectors though it is not discussed in this work. Persistence is an afterglow remaining on a device from a previous exposure of a bright source thought to occur due to a charge trapping mechanism in the detector material. See Smith et al. (2008); Deustua et al. (2010) and others for further discussion.

Reciprocity failure makes faint objects appear fainter than they really are. This has an obvious effect on Type Ia Supernovae (SNe) measurements planned by missions like the *Wide Field Infrared Survey Telescope* (*WFIRST*) as discussed in §1.8.2, biasing luminosity distance measurements. *Euclid* will use devices of this type for Baryon Acoustic Oscillation (BAO) and redshift space distortion (RSD) measurements where reciprocity failure, and its two dimensional structure, will affect BAO measurements by altering the spectral continuum as well as exaggerating absorption line depths. In addition, *Euclid* plans on using NIR detectors similar to the ones studied here to calibrate photometric redshifts, which may lead to increased noise and possible biases if they suffer from reciprocity failure. In the future HgCdTe devices may also be used for weak lensing measurements where reciprocity failure will alter the intrinsic galaxy shapes that must be recovered for accurate gravitational potential reconstruction. If the cause of reciprocity failure can be found, it may be possible to remove it. Until then, it is best dealt with by careful laboratory characterization of the effect. Devices with low reciprocity failure can be used for critical observations and the induced small biases can be subtracted off. In the absence of a sophisticated reciprocity failure setup, an estimate of the lower limit of reciprocity failure and its structure can be obtained by dividing flat field images, (properly masked and smoothed) taken at different illumination levels as in §5.4.5. On orbit measurements have also been performed (e.g. Bohlin et al. 2005; Riess 2010; Deustua et al. 2010) using sources such as white dwarfs. These do not match the accuracy and precision of laboratory based studies. Further cooling the devices also suppresses reciprocity failure. For low level reciprocity failure a modest drop in temperature may be enough to completely eliminate it. However, such cooling may be expensive, especially in space applications and there are hints that it may increase capacitive coupling in HgCdTe detectors as discussed in 6.2.5. This trade off would need to be studied carefully to determine the optimal operating temperature. There is also a possibility that reciprocity failure can be suppressed by increasing the voltage difference across the pixel readout MOSFET. This would be of help with low reciprocity devices if the voltage change can be kept small. A larger voltage change would adversely affect dark current. It is also not known how reciprocity failure may age or respond to radiation damage and further studies are necessary.

Subpixel response structure can also affect weak lensing shape reconstruction if HgCdTe devices are to be used in such measurements. It should be noted that CCDs may also exhibit structure that has so far escaped detection. The subpixel response appears to be a function of the source point spread function (PSF) size leading to complicated coupling between galaxy shapes and the subpixel structure. In addition to affecting shapes we found that the subpixel structure depends on the integrated signal level. This can significantly affect the recovered SNe magnitudes since the PSF that would be fitted to determine point source flux will not match the true *effective* PSF at high integrated signal levels. The subpixel structure may also affect BAO and RSD measurements in subtle ways. Measurements where the illumination does not change significantly over several pixels (like the black body spectral continuum) do not seem to be affected by subpixel structure. However, measurements of *thin* emission and absorption lines with thickness smaller than the width of a pixel may be affected and could lead to small wavelength offsets in these precision measurements. Depending on the resolution of its spectrograph, *Euclid* may need to account for this effect. The discovery of such complicated subpixel structure in HgCdTe detectors is a fairly recent development. Many more studies with a system such as the Spots-O-Matic are necessary to fully understand it. Until then it is best not to employ these detectors in an undersampled fashion.

I believe that the detailed measurement of QE described in Chapter 6 demonstrates a good starting point for a comprehensive detector characterization since it explores the response of the detector under various operating (temperature, voltage) and observational (flux, exposure time) conditions. Such a detailed measurement can reveal reciprocity failure and charge integration non-linearity and ways to mitigate them. A more complicated system such as the Spots-O-Matic is necessary to explore subpixel structure in detail.

APPENDICES

APPENDIX A

Mis-Centering Bias Reduction In Stacked SZ Observations

The analysis of the systematic uncertainties involved in stacking Sunyaev–Zel'dovich (SZ) signal on optical galaxy cluster positions in Chapter 3 illustrates how mis-centering of the SZ signal can suppress the recovered signal. Furthermore, the simple example in §2.5.2 of Chapter 2 also shows that the signal proxy, in that case the size of the smoothing kernel used in the Matched Filter (MF), can also alter the observed signal amplitude. We decided to investigate a method of reducing the negative signal bias by integrating the SZ signal within an aperture instead of simply using the central MF pixel as the amplitude proxy. Figure 2.28 in Chapter 2 in fact shows that simply using a larger smoothing kernel could recover some of the lost signal albeit at the cost of reduced signal to noise (S/N) ratio. However, simply integrating the filtered maps in apertures around cluster positions does not work due to ringing caused by the effective high pass filtering involved in the MF. One could use aperture integration on raw observed maps but the S/N degradation there would be extreme. Here we investigate how to compromise the need for low bias due to mis-centering with high S/N ratio.

A.1 Aperture Integration Failure

Figure A.1 shows the stacked filtered SZ profile of medium mass halos at low redshift. The perfectly centered stacked profile (blue line) has a higher MF amplitude (profile center at distance of $0'$) than the mis-centered stacked profile. The highest signal from mis-centered halos gets distributed around the final stack instead of adding to the amplitude at the very center. That is why the mis-centered signal (green line in Figure A.1) is higher than the perfectly centered signal at a distance of about $3'$. In principle, all it should take to recover the total signal would be to integrate out

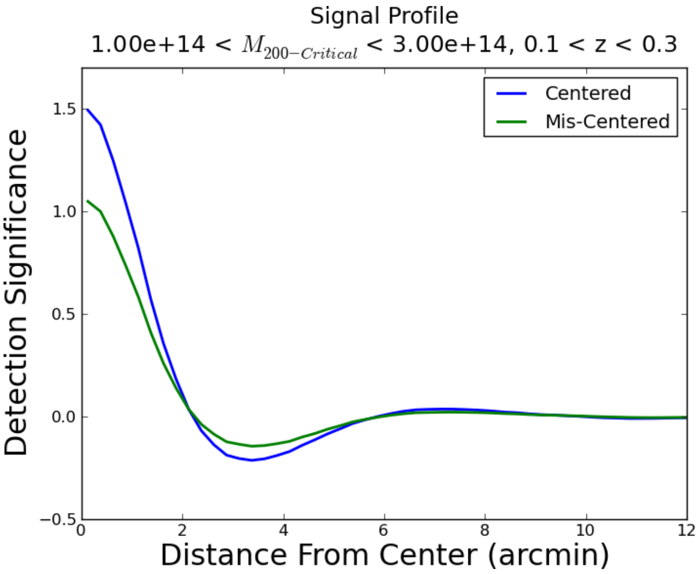


Figure A.1: Sample stacked centered and mis-centered radial profile shapes in a medium mass and low redshift bin. All single bin stacked data in this work use the same mass and redshift bin (labeled on the plot) for illustration.

to a radius beyond the likely mis-centering offsets. This would introduce more noise relative to the signal but would recover the total signal. This fails because the profile in Figure A.1 becomes negative between $2'$ and $5'$ distance from the center. Figure A.2 shows the result of such an integration over an aperture equal to the distance from profile center. Once the aperture size reaches the $2'$ radius the integrated signal begins to decrease for both the perfectly centered and the mis-centered stacks. In addition, Figures A.1 and A.2 show that fitting the profiles would not be a successful solution. The profile amplitudes are degenerate with the signal suppression due to mis-centering so there is no way to differentiate them. The cause of the negative signal in the profile is *ringing* caused by high-pass filtering introduced in the MF.

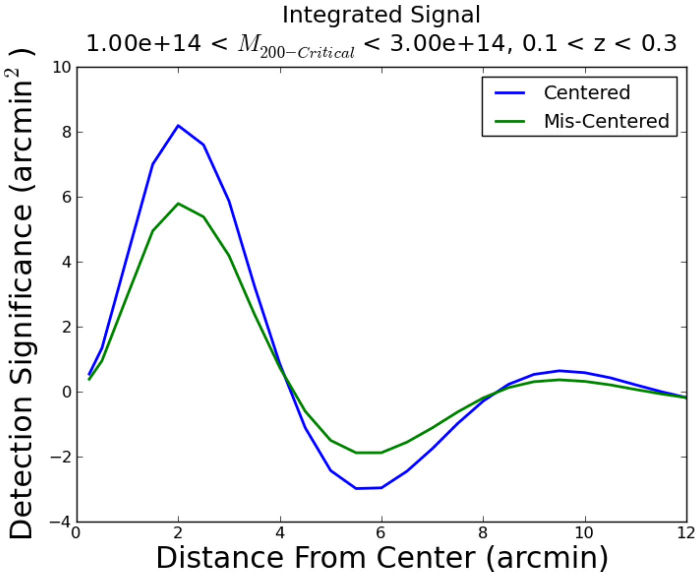


Figure A.2: Sample stacked and integrated centered and mis-centered radial profile shapes in a medium mass and low redshift bin. This is constructed by integrating the profiles in Figure A.1 cylindrically with element $2\,\pi\,r\,dr$.

A.1.1 Effective Filters

The Matched Filter (e.g., Herranz et al. 2002; Melin et al. 2006) is designed to extract the optimal S/N from either single or multi-frequency maps where the noise spectrum and signal profile shapes are known. In the case of *South Pole Telescope* (*SPT*) observations the primary noise contributions are white noise, the primary Cosmic Microwave Background (CMB) which contributes at low ℓ and is identical at all frequencies, isotropic atmospheric noise at low ℓ, scan direction $1/\ell$ noise, high ℓ instrument noise and unresolved point source contamination also at relatively high ℓ though it largely resembles white noise.

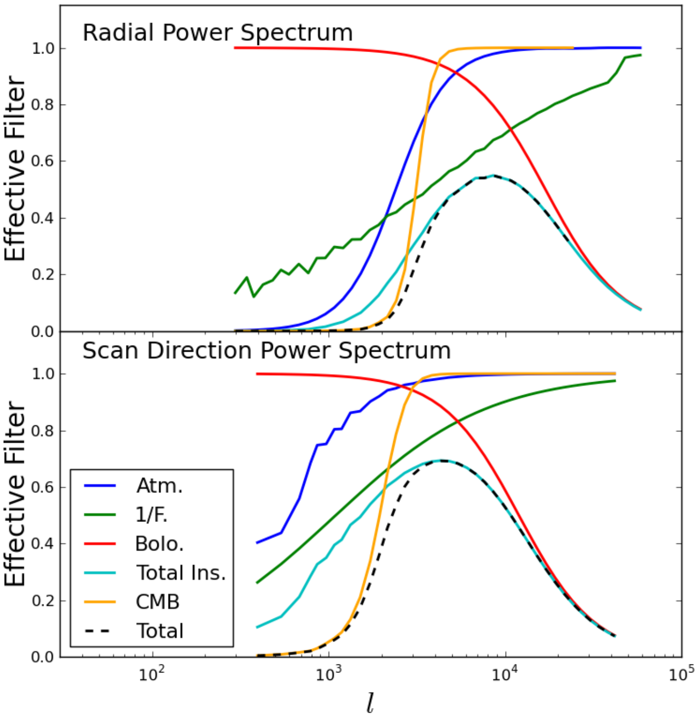


Figure A.3: Effective single frequency MF pass-bands. While the MF is not generally thought of in such terms, its application leads to such filtering in ℓ space.

The discussion of the undesirable effects of the MF, like ringing, is simpler in the language of signal processing. The MF is a combination of a high pass and a low pass filter. The *effective* filter employed by the MF is shown in Figure A.3 for the 150 GHz map though it is very similar for other *SPT* frequencies. The top panel shows the effective isotropic filter and the bottom panel shows the effective scan direction filter. In a single frequency implementation, the high pass filtering is provided in the isotropic direction by the inverses of the assumed atmospheric noise (blue line) and the primary CMB (orange line) spectra. Additional high pass filtering in the scan direction is due to the inverse of the $1/\ell$ spectrum (green line). The low pass filtering is caused by the inverse bolometer instrument noise spectrum in the scan direction (red line). The combined filter due to instrumental and atmospheric noise is shown as a cyan line and the total band pass filter that includes CMB is shown as a black dashed line. Finally, the map is smoothed with a kernel that approximately matches the shape of the cluster in SZ. This is also a low pass filter but is tuned to do the least amount of "damage" to the cluster signal while still helping to suppress small scale noise. In multi-frequency implementation of the MF, the primary CMB can be partially subtracted off, reducing or removing its contribution to filtering.

It should also be noted that *SPT* data is processed at a low level removing very low ℓ modes, $\ell < 300$ and with an exponential fall off of characteristic length of 318 in ℓ distance. Also, extremely high ℓ modes are removed. In our tests these do not impact signal recovery much and introduce a lot of statistical noise when deconvolved.

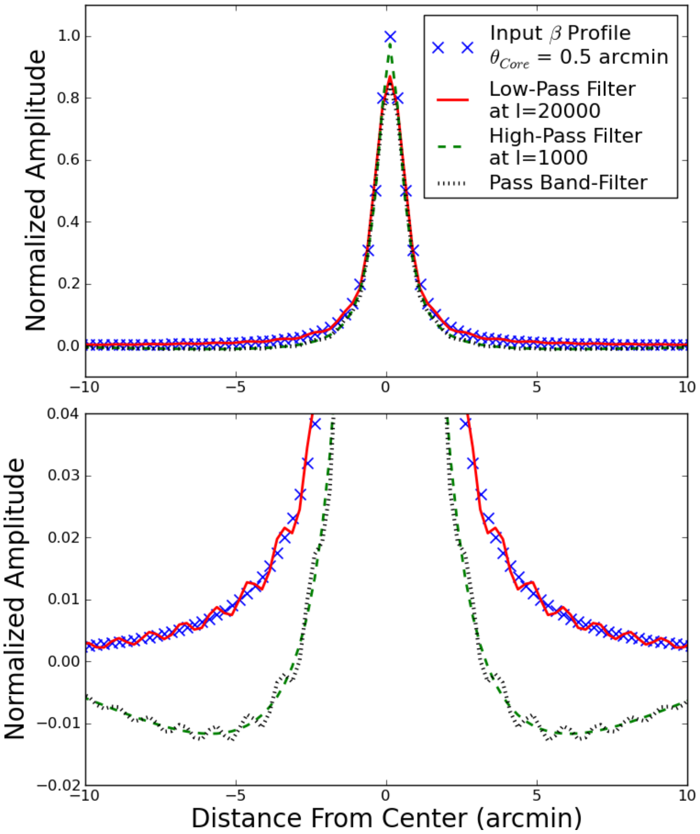


Figure A.4: A demonstration of ringing induced on a β profile by high, low and band pass filtering. Input profile is shown by blue × symbols. Impacts of high, low, band pass filters are shown by dashed green, solid red and dotted black lines, respectively. Bottom panel is a zoom in on the vertical axis of the top panel.

Figure A.4 demonstrates the impact of high (dashed green line), low (solid red line) and band (dotted black line) pass filtering on a β profile of half-arcminute size (blue × symbol). Low pass filtering introduces a small ringing at a relatively high frequency. High pass filtering causes large, negative dips around the profile. This ringing is then capable of *absorbing* the mis-centered signal when multiple profiles are stacked. We concluded that minimizing or removing the high pass components of the MF prior to aperture integration will reduce the bias caused by mis-centering. Of course, by reducing filter coverage we will also reduce the filter efficiency and increase statistical noise. Effectively, we will be trading the systematic bias in for increased statistical uncertainty. This may be desirable since a known statistical uncertainty is preferable to an unknown, possibly significant systematic bias. If the mis-centering bias however is characterized, this trade off may still be desirable if the uncertainty on bias is large.

A.2 Filter Tuning

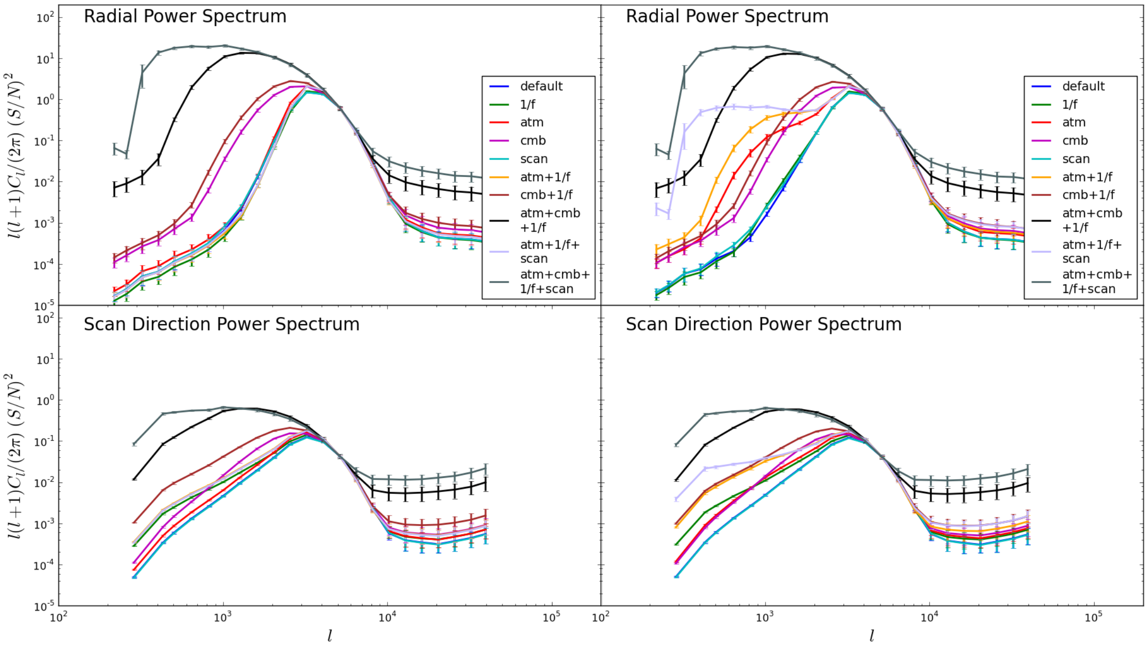


Figure A.5: The power spectra of recovered SZ signal after applying different variations of the MF. The spectra are normalized at $\ell = 5150$. The left panel shows the single, 150 GHz MF application. The right panel shows the multi-frequency, 150 GHz + 220 GHz MF application.

We now proceed to investigate the modifications of the MF. Figure A.3 is illustrative but it does not capture the full behavior of the MF, especially the multi-frequency implementation. Figure A.5 shows the results of various filtering schemes for a single frequency implementation in the left panel and for multi-frequency implementation in the right panel. It should be noted that only 150 GHz and 220 GHz simulated maps were available for this study. While the 220 GHz maps contain information on primary CMB they contain almost no SZ signal. 95 GHz maps should improve our results though they have yet to be included in this analysis. For this Figure a pure thermal Sunyaev–Zel'dovich (tSZ) map was convolved with *SPT* beams and then matched-filtered. No noise or backgrounds were included in the maps prior to filtering for clarity; we wanted to visualize the spectrum of the recovered signal in the absence of noise. Figure A.5 shows the results of matched filtering with various components turned off along with turning of the exponential fall off part of the *SPT* real-time scan-direction $1/\ell$ filtering as shown in the legend. These disabled components are, in order: none (default MF), $1/\ell$, atmospheric, CMB, *SPT* scan filter, atmospheric and $1/\ell$, CMB and $1/\ell$, atmospheric and CMB and $1/\ell$, atmospheric and $1/\ell$ and *SPT* scan filter, and atmospheric and CMB and $1/\ell$ and *SPT* scan filter. Note that the *SPT* scan filter was turned off by deconvolving the exponential fall off part of its profile from the data as it is nominally included in simulated raw images. Turning off the atmospheric, $1/\ell$, primary CMB and *SPT* scan filter (dark gray line) removes almost all high pass filtering with the only remaining component the low ℓ real-time *SPT* filtering. Of course, it will also introduce a large amount of noise and is merely a step above using raw data.

Of interest is the effect of atmospheric (red) and atmospheric and $1/\ell$ (orange) filter removal when compared across the left and right panels. The right panel shows that with these filters gone, a large fraction of the spectrum becomes accessible. This is not the case in the left panel. This illustrates that CMB filtering is already largely gone in the multi-frequency MF because it has optimally subtracted most of it while the single frequency MF has to remove it via a complete removal of information in the region of the spectrum the CMB occupies.

A.2.1 Filter Performance

We focused further on testing the atmospheric, CMB and $1/\ell$ filters. While the *SPT* scan filter deconvolution certainly helps extend the accessible ℓ range to lower values of ℓ, it does so at an extreme cost of noise. Figure A.6 shows the real space effects of the filters. The legend in the figure indicates which filters were *on* (unlike in Figure A.5). The single frequency (dashed lines) profiles have negative ringing unless the CMB filter is off. However, for the multi-frequency MF merely turning off the atmospheric filter (solid cyan line) removes the ringing though there is still some distortion and asymmetry due to the scan direction $1/\ell$ filter. In addition, turning the $1/\ell$ filter off further suppresses the S/N (green line) but makes the profile more consistent. This is further demonstrated in 2D using full simulation maps and filtering them in various ways. The top left panel of Figure A.7 shows the default MF results. Top right has the atmospheric filter turned off, the bottom left has the atmospheric

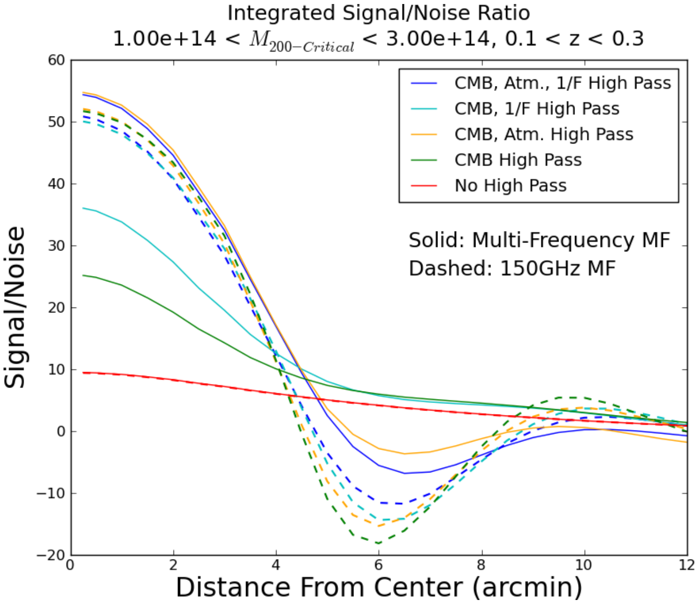


Figure A.6: Real space recovered SZ signal after applying different variations of the MF. Tests of atmospheric, $1/\ell$ and primary CMB filters are shown at single, 150 GHz frequency (dashed lines) and multi, 150 GHz + 220 GHz frequency (solid lines).

and $1/\ell$ filters off and the bottom right has all, atmospheric, $1/\ell$ and primary CMB filters off. This figure is for the multi-frequency (150 GHz and 220 GHz) while Figure A.8 shows the same filter types for a 150 GHz MF only. The single frequency images show negative ringing around the SZ halo unless all of the high pass filters are off. The multi-frequency images show small negative side lobes in the top right panel (atmospheric filter off) but no ringing in the bottom left panel, unlike in the single frequency implementation. Since there is no negative ringing, there is no place for the signal to get lost due to mis-centering.

A.2.2 Filters and Apertures

Figure A.9 shows the fractional uncertainty and bias (relative to perfectly centered halos) in the recovered Y_{SZ} from the fitted $Y_{SZ} - Mass$ relation in each redshift bin (rows in figure) as a function of various high pass filter configurations (columns) and smoothing kernel sizes (colors). The four columns show the default MF, with atmospheric off, atmospheric and $1/\ell$ off and all off, respectively. Shaded regions show the statistical uncertainty vs integrating aperture size while the lines show the systematic biases vs integrating aperture size. The obvious spikes in the statical uncertainties and in biases are caused by the signal (in this noise to signal ratio) going to 0 due to ringing caused by the high pass filtering. This ringing occurs with the 3 smallest kernel smoothing sizes for the default matched filter but disappears for all but the smallest smoothing kernel in the rest of the filter implementations. The goal for us is to minimize the systematic biases without allowing the statistical uncertainty to

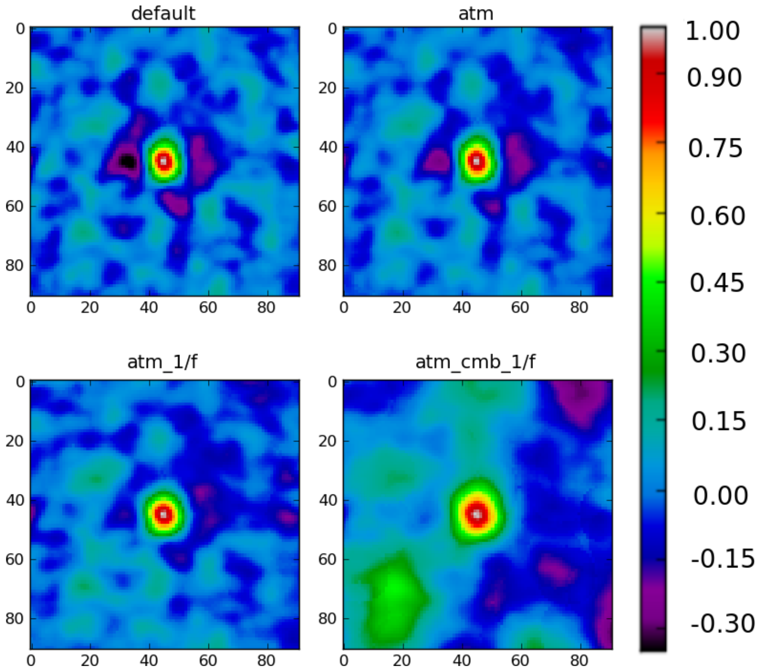


Figure A.7: Stacked, 2D, matched-filtered SZ signal using multi-frequency MF. Panel labels indicate the filters *turned off* for the plot. Stacks are normalized to peak signal. See text for details.

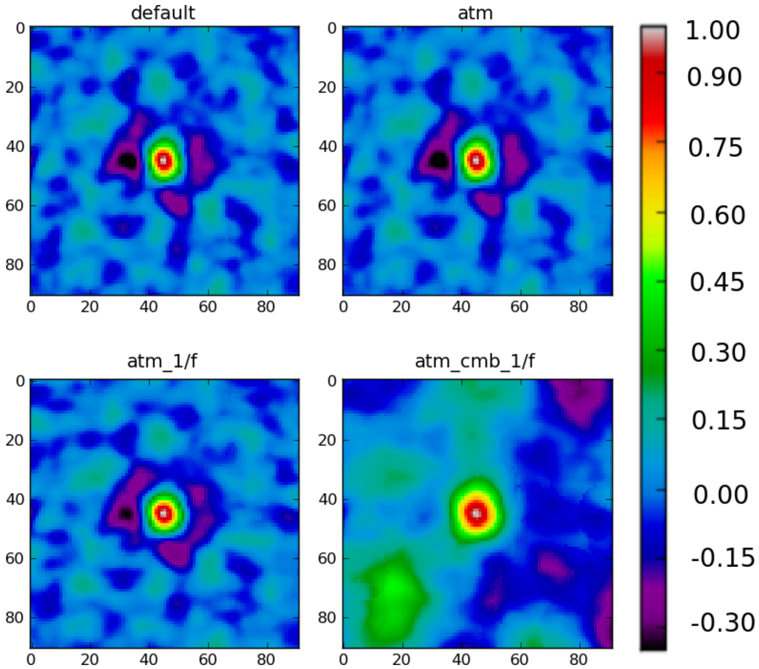


Figure A.8: Stacked, 2D, matched-filtered SZ signal using single frequency MF. Panel labels indicate the filters *turned off* for the plot. Stacks are normalized to peak signal. See text for details.

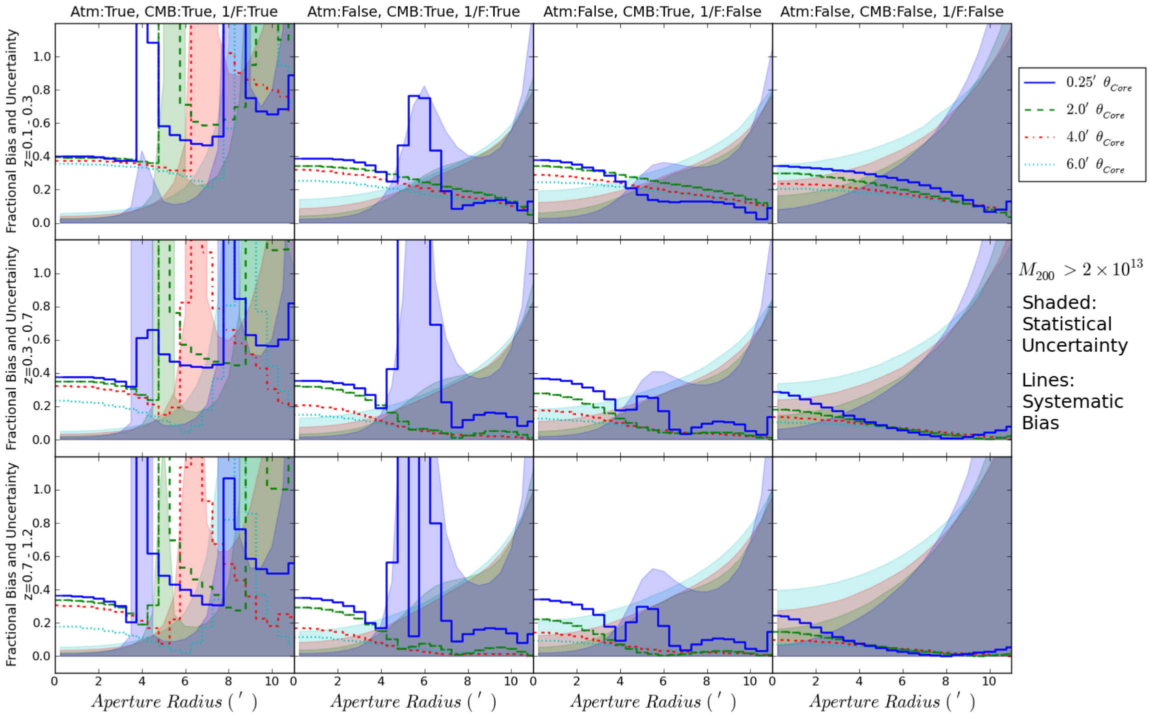


Figure A.9: Residual (relative to perfectly centered) uncertainties (shaded) and systematics (lines) for various combinations of high pass filtering (columns) in redshift bins (rows) average over masses larger than $2 \times 10^{13} h^{-1}\,M_{\odot}$. Colors indicate different smoothing kernel sizes. See text for details.

blow up. For instance, with all high pass filtering off in the right-most column, a sufficiently large aperture (that is redshift dependent due to angular diameter scaling) completely removes any bias. However, the statistical noise there is enormous. In addition to this comparison, Figure A.10 shows the same information with the systematic bias relative to the default mis-centering model. That is, the systematic here is based on the uncertainties in the Johnston et al. (2007) mis-centering model while the mean mis-centering gives a bias of zero. To summarize, the previous Figure tells us what the biases are when we completely ignore mis-centering while this figure tells you what the biases will be if the mis-centering distribution is somewhat characterized (there is a prior on it) and included in our analysis.

A.3 Results

As indicated above, there are a lot of parameters that can be tuned to suppress the mis-centering bias. We further discuss some particular choices. We limit ourselves to

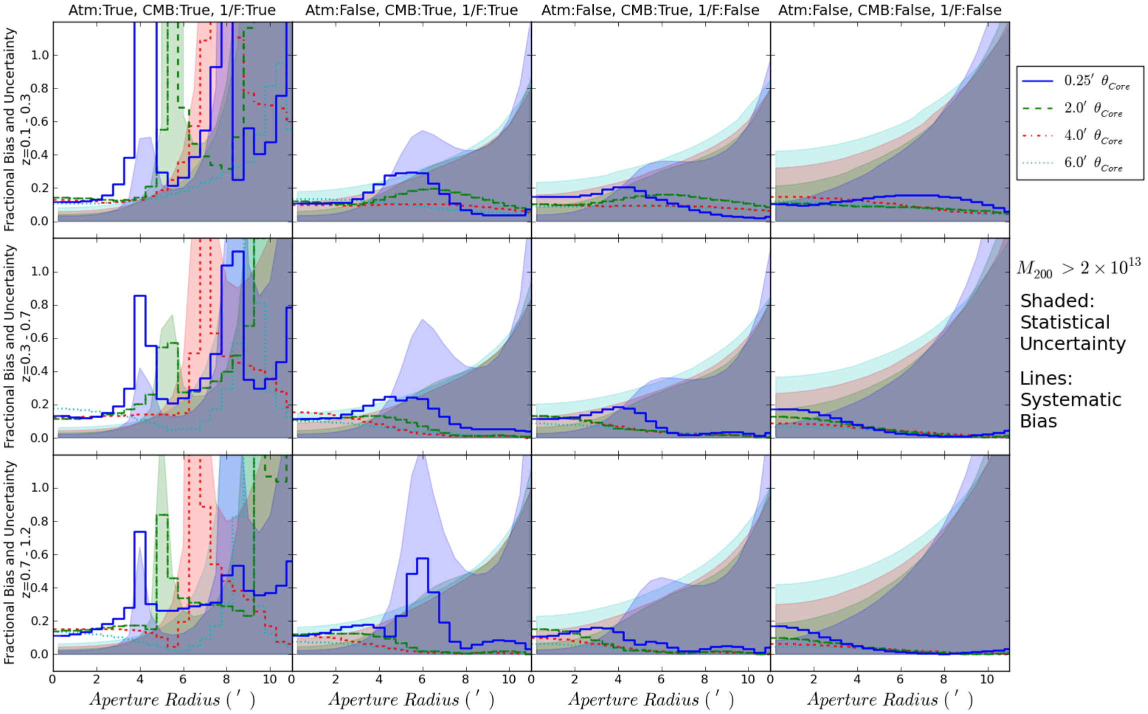


Figure A.10: Residual (relative to default mis-centered model) uncertainties (shaded) and systematics (lines) for various combinations of high pass filtering (columns) in redshift bins (rows) average over masses larger than $2 \times 10^{13} h^{-1}\, M_{\odot}$. Colors indicate different smoothing kernel sizes. See text for details.

the multi-frequency implementation of the MF due to its clear superiority in primary CMB suppression. In addition, I will now hold the smoothing kernel fixed at $2'$ since it was sufficient to suppress divide-by-zero errors in the three right-most columns in Figures A.9 and A.10. And I will discuss two aperture sizes, $4'$ and $6'$ since that is a parameter that will likely change depending on the redshift bin under consideration when stacking. We'll first look at the case where the biases come from completely ignoring mis-centering uncertainty in §A.3.1 followed by the case where they come from the uncertainty in the assumed mis-centering prior in §A.3.2.

A.3.1 No Prior On Mis-Centering Distribution

Figure A.11 shows the recovered signal proxy (black stars) and statistical uncertainty (black shaded region) for the perfectly centered simulation set and the recovered signal (blue circles), statistical (blue shaded region) and systematic (blue lines) uncertainties for the mis-centered model from Johnston et al. (2007) in three redshift bins (rows) and three MF implementations (columns). More detail is visible when the residual of the centered and mis-centered models are plotted against the perfect expectation in Figure A.12. The default MF signal (left panels) differs significantly between the statistically allowed range (gray bands) and the allowed mis-centered systematics (blue lines). Reducing the high pass cutoff frequency and increasing the integration aperture reduces the extent of the bias at the cost of increasing statistical noise (gray and blue bands). The blue and black shaded region at low redshift overlap significantly in the top right panel and at mid redshift in the center panel. At those settings, the residual bias is smaller than with the default MF and dominated by the statistical uncertainties. Hence, any cosmological applications of SZ measurements will likely benefit from a modified filtering scheme.

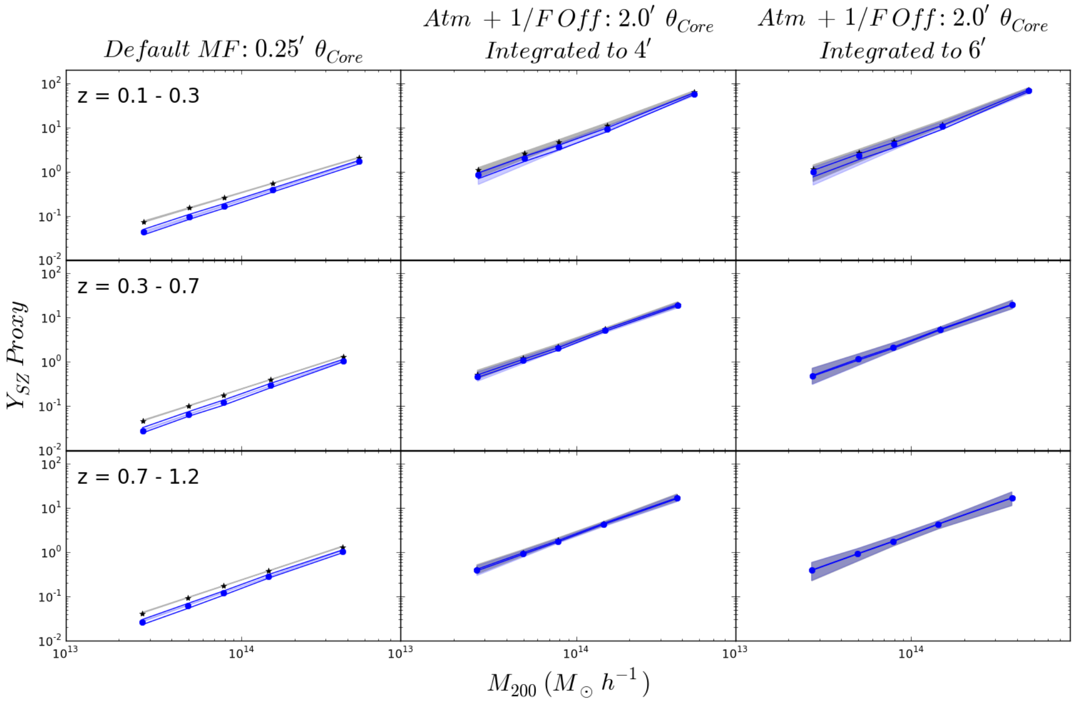


Figure A.11: Y_{SZ} signal proxy vs mass. The 3 columns show the signals obtained using the default matched filter, high pass reduced + integrated over $4'$ and high pass reduced + integrated over $6'$ matched filters, respectively. The 3 rows show the redshift bins used. The black stars show the simulation scaling for perfectly centered halos while blue circles show the scaling for the Johnston et al. (2007) mis-centering model. Shaded bands of both colors are shown though details are difficult to discern here. See Figure A.12 for a more informative plot.

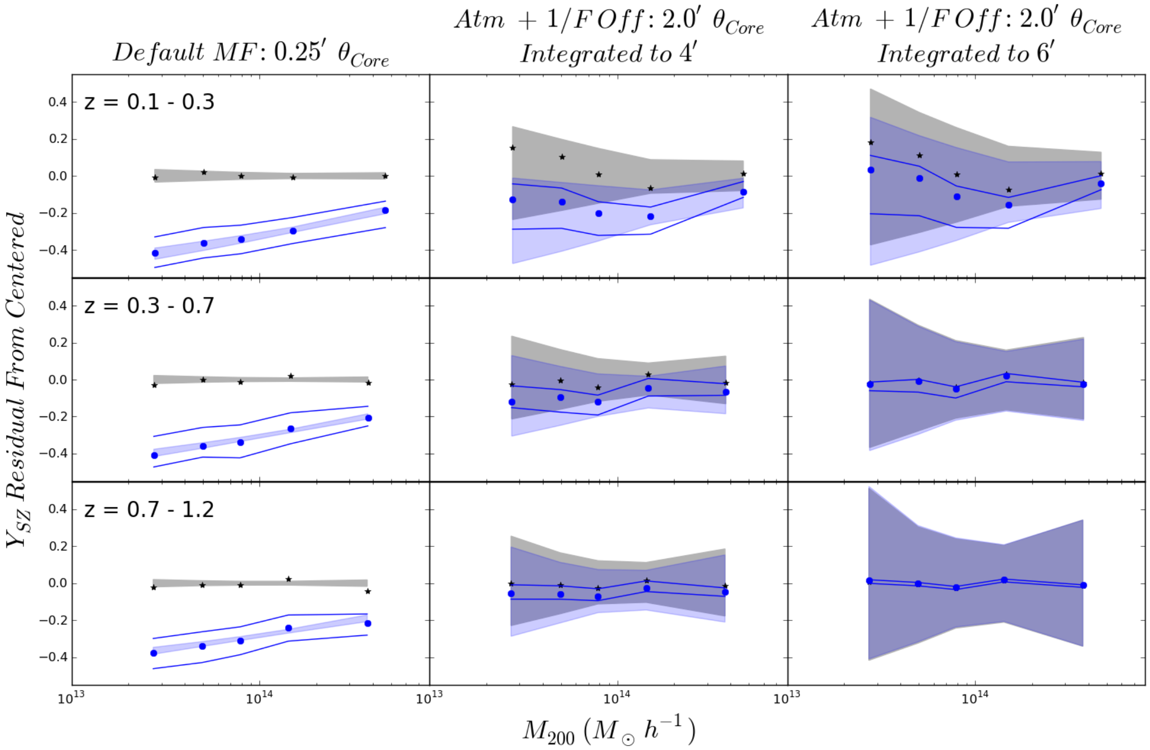


Figure A.12: Fractional Residual Y_{SZ} signal proxy relative to the signal obtained from the perfectly centered distribution. The 3 columns show the signals obtained using the default matched filter, high pass reduced + integrated over 4′ and high pass reduced + integrated over 6′ matched filters, respectively. The 3 rows show the redshift bins used. The black stars show the simulation scaling for perfectly centered halos while blue circles show the scaling for the Johnston et al. (2007) mis-centering model. Shaded bands of both colors show the statistical uncertainties. The solid blue lines show the $1\,\sigma$ bounds on the mis-centering model from Johnston et al. (2007).

A.3.2 With a Prior On Mis-Centering Distribution

Figure A.13 also shows the residual deviation of the mis-centered model but it is calculated against the default mis-centering model instead of the perfectly centered one (note the lack of black stars and black shaded region). This case may be a more realistic application where the mis-centering distribution is known a priory to some degree and any bias is due to the uncertainty (solid blue lines) on that distribution. Again, modifying the filter does reduce the systematic uncertainty due to mis-centering but the statistical uncertainty introduced tends to be much larger than the original systematic. It may be preferable to propagate these systematics into final cosmological measurements instead of employing our MF modification scheme. It is possible that the 95 GHz data would reduce the statistical uncertainties enough to make this approach preferable.

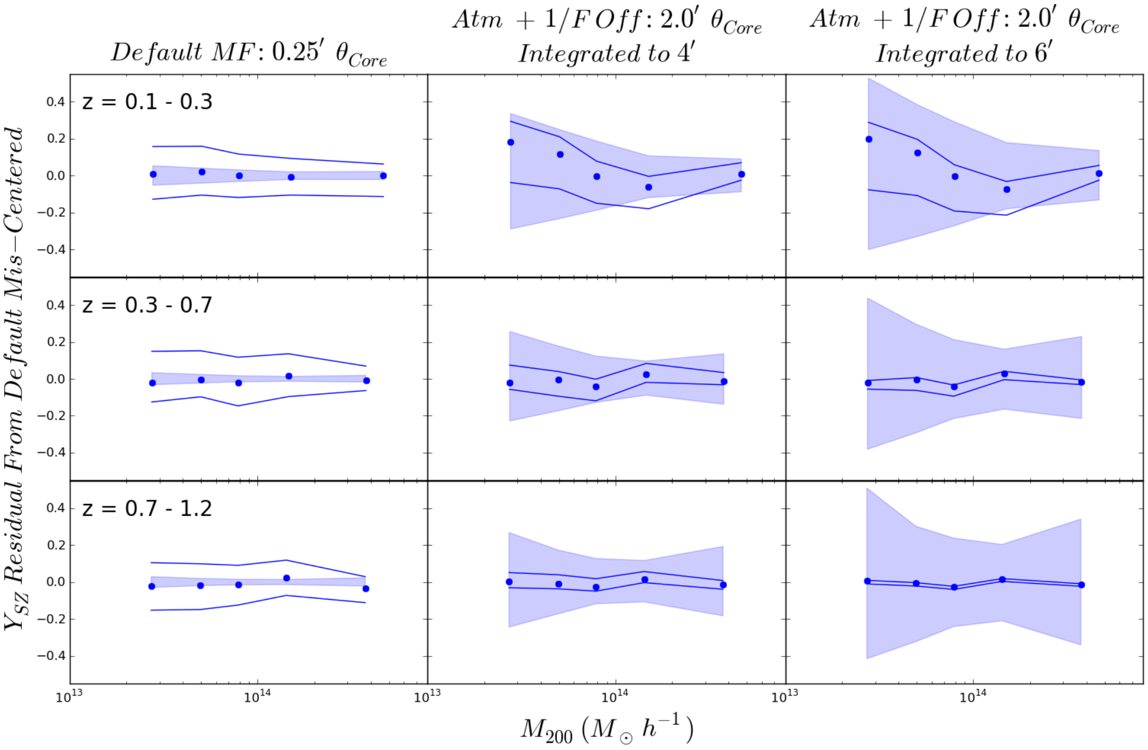


Figure A.13: Fractional Residual Y_{SZ} signal proxy relative to the signal obtained from the mean mis-centering distribution. The 3 columns show the signals obtained using the default matched filter and the 3 rows show the redshift bins used (see caption of Figure A.12). While the default MF signal (left panels) is now consistent with zero, the $1\,\sigma$ bands are substantially larger than the statistical uncertainties (blue lines and blue shaded region, respectively).

A.4 Discussion

Reduction in mis-centering – induced bias requires two steps, high frequency filter reduction and aperture integration. Removing the primary CMB, atmospheric and $1/\ell$ high pass filters leads to the most robust resistance to mis-centering bias. With sufficiently large integration radius all signal will be recovered even when very small smoothing kernels are used. This applies to both multi and single frequency matched filters. The multi-frequency matched filter allows for a compromise where the atmospheric and $1/\ell$ high pass filters are turned off with the remaining primary CMB filter suppressing low frequency modes far less than it would in a single frequency MF.

The ability to tune filter performance is important since re-introducing low frequency modes significantly increases statistical noise.

Note that further deconvolution of the the high pass filter applied by *SPT* (with a cutoff at $\ell = 300$ and an exponential fall off with characteristic length $\ell = 318$) serves only to increase statistical noise since the modes re-introduced in this process fall far outside any reasonable integration radius. We have experimented with introducing additional high pass filters with tuned cutoffs to optimize the noise and bias performance. We saw no or only minor improvements but with a more thorough investigation it may be possible to fine tune such a filter.

Projection effects in SZ measurements will get worse due to integration but they can be subtracted off with only minor noise contributions by integrating and stacking around many *empty* regions of the millimeter-wave sky.

It is likely that the addition of the 95 GHz data will improve our results since in helps to constrain the primary CMB while contributing SZ signal. Figure 2.23 in Chapter 2 does show that MF cluster–finding improves when utilizing all three frequencies, though a similar plot with only 150 GHz and 220 GHz does not show an improvement over a 150 GHz cluster–finder alone.

APPENDIX B

Statistical Investigation of Response Structure in a NIR Detector

Our goal is to understand the detector response structure of device H2RG-103. That means that we must know what this structure looks like and what physically leads to it. We believe that there are several contributing effects:

1. Large scale sensitivity variations due to the bulk or surface properties of the HgCdTe material and/or its coating. These can correlate pixels over dozens to hundreds of pixels and are not correlated with the readout channels.

2. Readout sensitivity variations due to how the device is read out. Individual channels (there are 32 of them, each one reading out 64 adjacent columns of the device) have different gains due to bias voltage differences. While those have been removed as best as we can, residual differences can remain. In addition, there are hints of gradients along the rows in readout channels seen in HgCdTe devices though this is not very prominent in H2RG-103. In addition, since each channel is read out via the same analog and digital circuits other sensitivity correlations may exist.

3. Small scale sensitivity variations. These occur due to differences in the electronic characteristics of individual pixels and are not spatially correlated though they do form an irreducible noise floor.

4. Pixel size variations. Since the area of a detector is conserved, if a larger than average pixel active area extends into the area nominally covered by an adjacent one, that pixel's effective area decreases.

5. Line features. The source of these is not fully understood. Some of them are likely sensitivity variations due to surface characteristics of the HgCdTe

material or possible stresses in the bulk. Others may arise due to offsets in photo-lithography mask placement when the pixel grid is laid out which may increase or decrease the area of every nth column or row, depending on the mask placement direction.

The great majority of the analysis discussed here is based on a small, 320 × 320 pixel section of the full detector (2048 × 2048). This device suffers from large sensitivity variations and this section was the largest one that fulfilled several requirements. The overall gradient across the section was small and there were no large masked regions with preferred directionally. There are a few clusters of bad pixels but they are fairly random and should therefore not have too much of an effect on our analysis tools. The entire 2048 × 2048 pixel detector was used only once for the correlation function shown in Figure B.3 where increased statistics were needed to show an effect.

B.1 Background

It was observed (Lorenzon et al. 2008) that the spatial noise at small scales present in HgCdTe detectors, specifically device H2RG-103, was not due to purely photon counting statistics, that is, shot noise. Averaging of multiple exposures of a device should in principle lower the noise in the average image proportionally to $1/\sqrt{N}$ where N is the number of samples included in the average. Figure B.1 shows that this is not the case since the measured noise, blue ×, reach a systematic floor far above the expectations based purely on shot noise (black line). The 64 × 64 pixel squares used for the above analysis reached a noise floor of 1.05% when one hundred of them were average. They are not completely free of large scale structure. As Figure

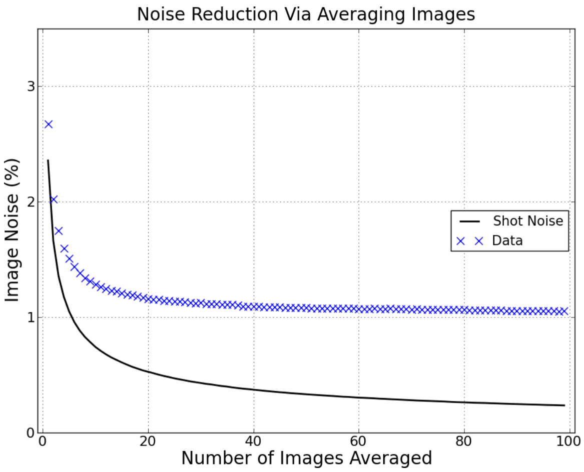


Figure B.1: Noise reduction in the average of multiple flat-field exposures. The 320 × 320 pixel *flat* section of the device was split into twenty five 64 × 64 pixel squares. Each square was sigma-clipped and its standard deviation was computed. The median noise was chosen from the 25 sections and plotted here as blue ×. The same procedure was applied to a random map (with capacitive coupling applied) and its results are shown as the black line.

B.7 will show there is still a Large Scale Structure (LSS) contribution to the noise coming from detector response gradients. Using smaller sections, 15 × 15 pixels, Lorenzon et al. (2008) have shown that a slightly smaller noise floor of 0.83% can be reached. Nevertheless, the larger image regions are necessary to perform the tests described here and though the LSS contribution is not significant at small scales it is still modeled in this work.

The fact that the spatial noise of many averaged exposures reaches a systematic floor indicates a presence of an irreducible noise floor. This noise floor can be caused either by very small scale (≈ one pixel in size) sensitivity variations, pixel size variations previously observed in CCD's (Smith & Rahmer 2008), or a combination of both. In addition, a variety of larger scale line-shaped variations called "scratches" (though that may not be the actual cause) have been observed adding to the overall noise floor.

B.2 Detector Response Characterization

Several statistical tools have been developed to distinguish between the various sources of noise and to determine their relative contributions to the total irreducible noise background of device H2RG-103.

B.2.1 Correlation Function Analysis

The purpose of the two-dimensional correlation function is to describe how likely it is on average that two pixels, separated by fixed horizontal and vertical distances have the same value. This can help distinguish between negatively correlated pixel size variations and positively correlated sensitivity variations. It can also visualize various line features. The correlation functions are normalized to the image variance meaning that the values of the correlations range between +1 (complete correlation) to -1 (complete anti-correlation). A value of zero indicates a lack of correlation. The correlation function is defined as:

$$c(i,j) = \frac{1}{\sigma^2} \sum_{x=0}^{N_x-i} \sum_{y=0}^{N_y-j} \frac{(s_{xy} - \langle s \rangle)(s_{(x+i)(y+j)} - \langle s \rangle)}{(N_x - i)(N_y - j)} \ , \tag{B.1}$$

where σ^2 is the variance and $\langle s \rangle$ is the mean of the region over which the correlation function is being computed, N_x is the number of columns, N_y is the number of rows and s is the value of a given pixel. In principle the correlation function describes the device behavior completely however in realistic devices, large scale structure makes this method intrinsically scale dependent since it is computed based on the value of a pixel relative to the mean value of the region under consideration. Therefore, extraction of quantitative results is best accomplished by comparison to simulations. The two dimensional correlation function nevertheless provides valuable qualitative as well as order of magnitude guidance.

The two-dimensional correlation function produces a map at a given scale. For example, a map with length scale of 64 pixels has 129 by 129 values ($2 \times [64] + 1 = 129$). It defines the correlation between pixels that are separated by 0,1,...,64 pixels left/right and up/down directions. The center of the map has the correlation at a vertical and horizontal separation of zero. It is in fact the variance of the data and, because of the normalization, always has a value of one.

Figure B.2 shows the correlation function at length scale of 16 pixels (top left panel) and 64 pixels (top right panel). The bottom panels zoom into the center region of the top panels. In practice, the flat section used for this analysis was clipped to remove outliers and high pass filtered, using a simple RC type filter, with a characteristic filter scale of 16 pixels for the left panels and 64 pixels for the right panels.

Several features are instantly obvious. Pixels separated horizontally by one pixel are negatively correlated. We interpret this as likely due to pixel size variations. However, this correlation does not exceed -15% (in left panel) which means that it cannot be solely responsible for the irreducible noise floor of the device; in such a case the correlation would be of order 50%.

In addition, there appear to be 3 linear structures, strongest at 18^o from horizontal, mid-strength one at about 86^o from horizontal and one at -50^o from horizontal. There is also a hint of a weak vertical correlation. It may be related to a positive correlation every eight column more easily visible in the correlation function computed over the entire device due to the better statistics. It is shown in Figure B.3. This figure also shows a diffuse correlation at 45^o that is likely produced by the bottom left corner of detector H2RG-103 which is not covered by anti-reflective coating with a boundary at 45^o. Thanks to the correlation function analysis we now have an idea of the parametrization necessary to describe the device response. There is large scale structure (visible if high pass filtering in the correlation function computation is off or too low). There is also horizontal negative correlation between adjacent pixels that we interpret as pixel size variations. There are linear features present and it is also likely that there is an irreducible random background due to individual pixel

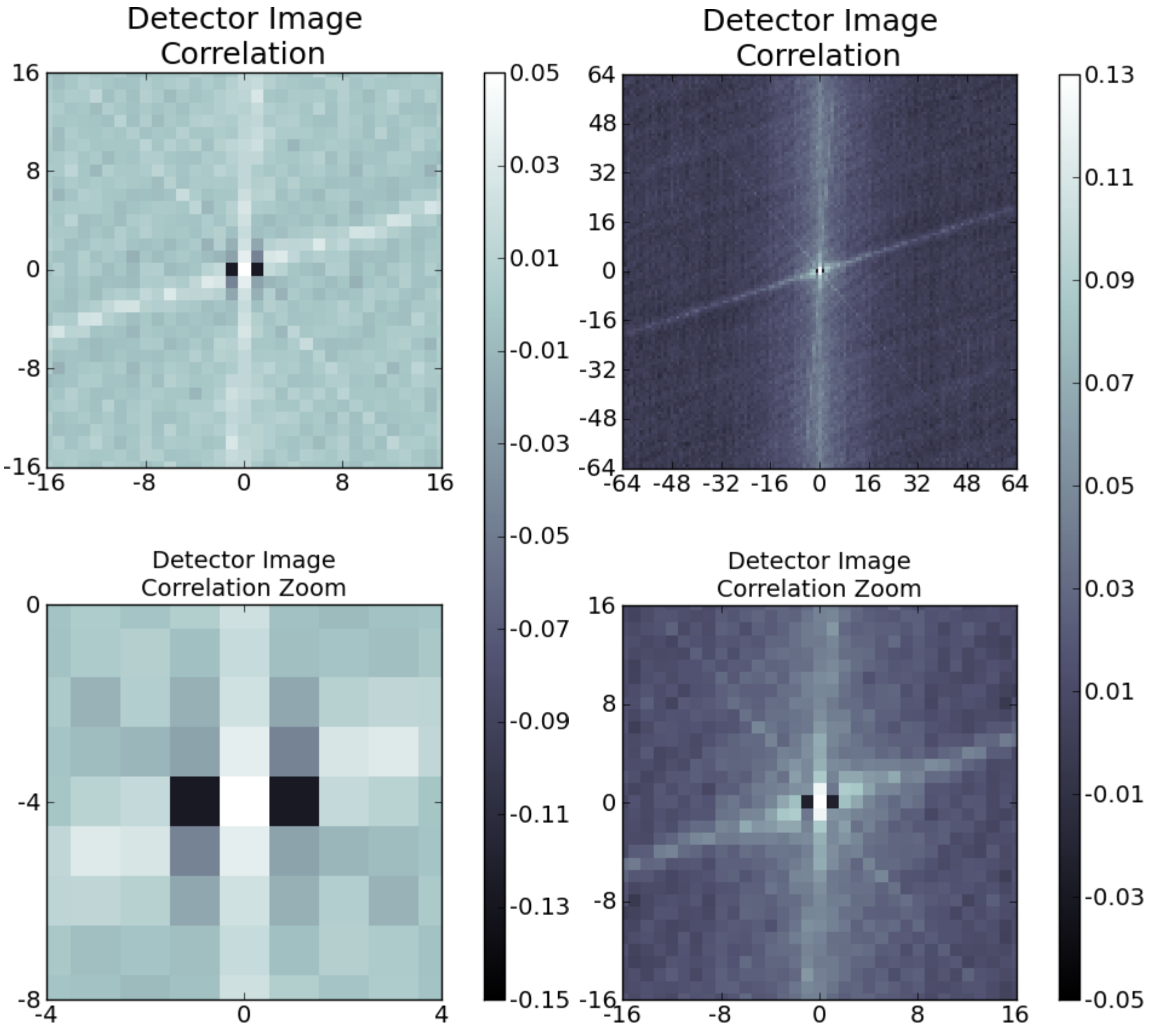


Figure B.2: Correlation functions of pixels up to 16 pixels apart on the left and up to 64 pixels apart on the right. Bottom panels are zoomed in version of the top ones.

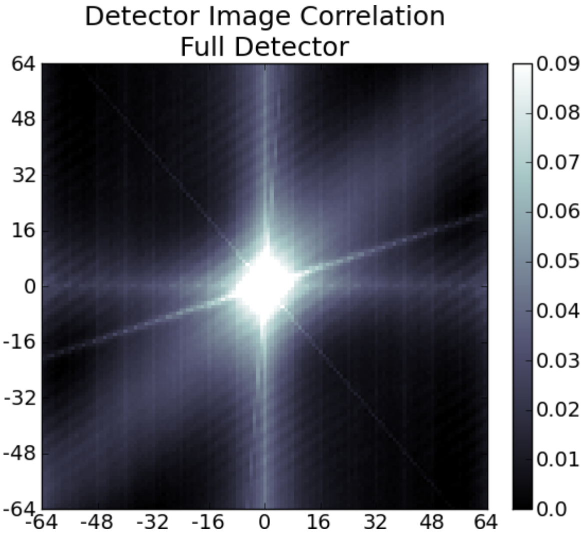


Figure B.3: Correlation function of pixels up to 64 pixels apart. This one is computed over the entire device instead of the *flat* region used for most of the analysis described here. The color scale is enhanced to distinguish the eight column correlation hence the center of the figure is saturated.

electronic characteristics since the adjacent pixel correlation cannot be responsible for the noise alone.

B.2.2 Angular Noise Spectrum

The angular noise spectrum method is designed to indicated whether the noise distributions along lines (at some angle relative to horizontal) deviate from Gaussianity. It is particularly good at detecting and quantifying linear features in the detector response.

At a given angle from horizontal, anything between -90^o to 90^o in steps of 1^o, all of the pixels are assigned to some line at that angle. There are many such lines. The means of all such lines are computed. The variance of those means is then multiplied by the length of the lines to obtain a prediction for the variance of the pixels along these lines assuming their values are normally distributed. By comparing the expected noise, that is $\sqrt{variance}$, of the pixel values at some angle to the noise of the image, correlations can be identified. If negatively correlated pixels are located along a given line the noise of the pixels computed from the variance of the means will be lower than the overall image noise. Sensitivity correlations will lead to a higher noise. In principle, this method could be used to look for any desired feature instead of just straight lines by computing the variance of the means of pixels along an arbitrary path.

Figure B.4 shows the angular noise spectrum of the flat detector region at three different scales: squares of 32, 64 and 128 pixels on each side. They show somewhat different things as certain noise sources can be more prominent over smaller or larger scales. This spectrum shows the same line features that are easily observed in the correlation function in Figure B.2. The 18^o linear feature is very prominent at scales of 64 and 128 pixels while the one at about 86^o from horizontal and the vertical one become strongest at 128 pixel scales indicating that they have a very long correlation length. The linear feature at -50^o is strongest at scales of 64 pixels. In addition, the gentle dip around zero degrees may be indicative of adjacent pixel size variations.

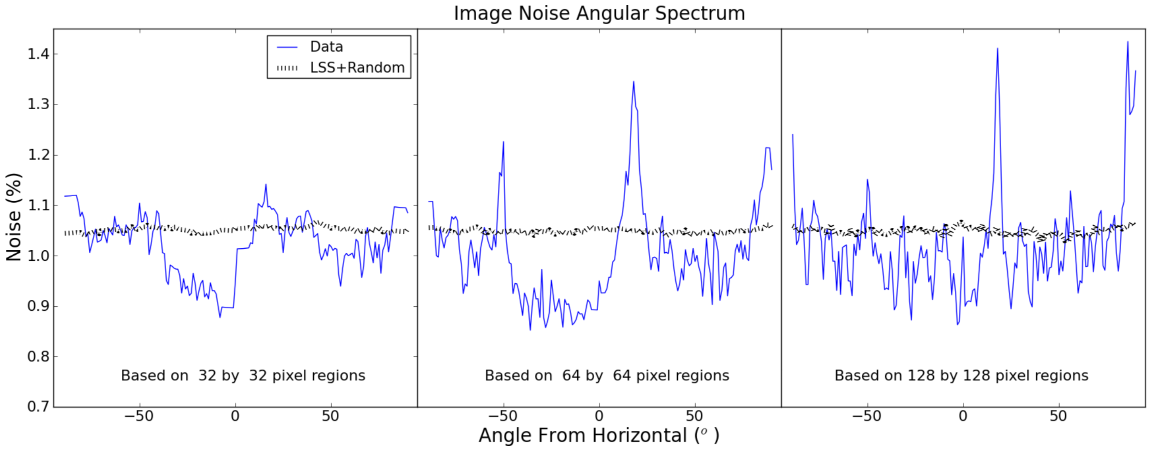


Figure B.4: Angular noise spectrum. The flat detector image was divided into smaller squares. The angular noise spectrum was computed for each square and averaged for the entire flat image. These squares were 32 × 32 pixels wide in the left panel, 64 × 64 pixels wide in the middle panel and 128 × 128 pixels wide in the right panel.

B.2.3 Multiple Exposures Averaging

Observing the noise reduction in an averaged image as more and more frames are included in the average is probably the simplest analysis technique. The blue × symbols in Figure B.5 show the same data as in Figure B.1 but in a log-log scale. What makes this more useful is the addition of vertically and horizontally partial-averaged individual images which are then averaged over multiple exposures.

For vertical partial averaging, every 4 pixels in a given column are averaged into a single one, turning our analysis region from 320 pixels by 320 pixels into a region of 80 pixels by 320 pixels. For horizontal partial averaging every 4 pixels in a row are treated as such leading to a 320 pixel by 80 pixel single exposure frame. For a spatially random background, the noise in these partially-averaged images averaged over exposures should scale as $1/\sqrt{N}$ just like the full image averaging but be a factor of $\sqrt{4} = 2$ smaller than it. This is shown by the black solid, dashed and dotted lines in Figure B.5. However, for pixels with positive correlation along either the vertical or horizontal direction the reduction in noise will be less than a factor of two and for pixels with negative correlation the reduction will be more than a factor of two. This is demonstrated by the red lines in the figure. I generated a model where the entire irreducible noise background came from horizontal pixel size variation. The solid red line reaches the same systematic noise floor as the data since that is what I put in. However, the horizontally partially-averaged frames, dotted red line, have systematic noise far smaller than just a factor of two reduction would provide as the vertically partially-averaged frames in dashed red demonstrate. In the real data, both the vertically (dashed blue) and horizontally (dotted blue) partially-averaged frames do not reach a factor of two noise reduction over the overall (blue × symbols) indicating the presence of positive correlations in vertical and horizontal directions. However, horizontally (dotted blue) partially-averaged frames do show a larger decrease in noise which is consistent with the presence of small-scale anti-correlations in that direction such as pixel size variation.

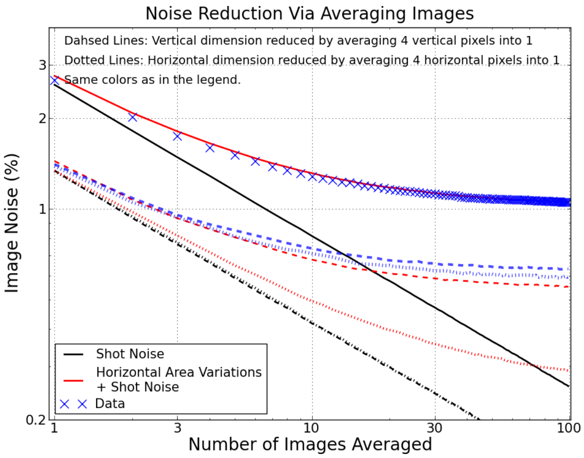


Figure B.5: Noise reduction. Similar to Figure B.1 but shown on log-log axis to emphasize small details. In addition to the overall noise reduction (blue ×), partially-averaging individual exposures in vertical (dashed) and horizontal (dotted) directions are shown. Predictions based on shot noise are shown as black solid, dashed and dotted lines and predictions based on an irreducible horizontal pixel size variation are shown as red solid, dashed and dotted lines.

B.3 Modeling H2RG-103

In order to test our understanding of the irreducible noise sources present in detector H2RG-103 I simulated images to attempt to reproduce the gross features of the device. I decided to model only the most important and interesting contributions: large scale sensitivity variations, the 18^o degree line, pixel area variations and irreducible random noise. This is accomplished in three separate steps in addition to applying a known capacitive coupling factor (Moore et al. 2004; Brown et al. 2006; Barron et al. 2007).

First, I manually adjust the strength of the line feature. Then, I fit the low k modes of the isotropic power spectrum with a power law to obtain the large scale sensitivity variations. Finally, the average isotropic and one dimensional power spectra computed over a small square region are fitted to derive the strength of the pixel area variations and random irreducible noise.

To complete a full simulation, random shot noise is added to 100 images with the same underlying irreducible response structure followed by capacitive coupling. The 100 images are then averaged to allow for a direct comparison to the data.

B.3.1 18^o Line Feature

Figure B.6 shows the angular noise spectra of the data (blue line) and the 18^o degree line model (orange dashed line). They have been smoothed with a boxcar kernel with width of 5^o. The line is simulated as stretching over the entire flat region.

At each row, the corresponding column coordinates are computed based on the angle of the line drawn from a Gaussian distribution centered at 18^o with σ=3^o. That line then gets an amplitude assigned from a half-Gaussian distribution; random draws larger than zero are thrown out to force the amplitudes to be negative since the easily visible line feature tends to be darker than the mean of the image. The mean of this line map is then rescaled to zero. The result is that there are a large number of lines with positive signal relative to image mean but the most biased lines are always negative. The amplitude was not fitted but rather adjusted manually to roughly match the observed angular noise spectrum; the σ of the half-Gaussian was set to 0.2% of the mean map level. This is by no means the optimal procedure. We simply do not know the true distribution of these features. The goal here is to reproduce the gross behavior of this particular sensitivity variation type. Note that capacitive coupling was included.

B.3.2 Large Scale Structure, Area Variation and Random Noise

The modeling of the large and small scale structure was somewhat more quantitative and was done by fitting the real noise power spectrum after accounting of the line feature above. Figure B.7 shows the overall (left panel) as well as the small scale (right panel) noise power along with best fits. The figure shows noise power spectra computed isotropically versus radial wave number as blue circles, as well as in one dimension, vertical as red diamonds and horizontal as red squares. The left panel shows the power spectrum computed using the entire 320 × 320 pixel flat detector

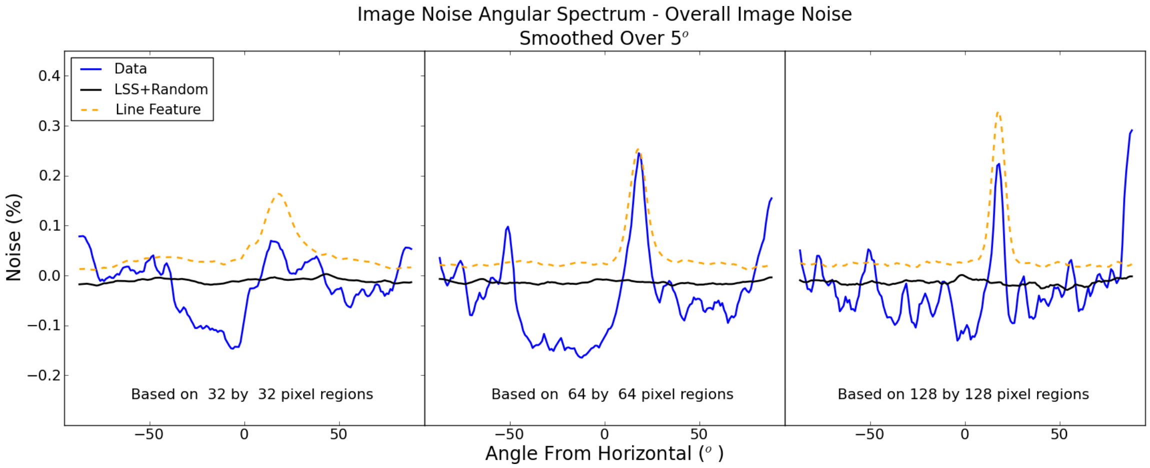


Figure B.6: Modeling angular noise spectrum at three scales. See the caption for Figure B.4 for details on the scales. The blue line shows the smoothed noise spectrum of the data. The dashed orange line shows the modeled linear feature noise spectrum. The black line is the expectation based on a random realization (with large scale variation added for completeness though it has no impact here).

image area using simple Fourier transforms. The right panel was somewhat more complicated to obtain. The flat area was subdivided into twenty five 64 × 64 pixel square regions. Power spectra were computed over each square and then averaged over the twenty five squares. This provides finer detail as well as lowering large scale contamination.

The spectrum of the full region is not surprising. The noise is mostly isotropic though the horizontal power is somewhat larger at the largest scales. However the small scale spectrum is very different. While the vertical power spectrum (green diamonds in the right panel of Figure B.7) reach a noise floor, the horizontal, and to a smaller extent the isotropic, power spectra begin to rise past wave number of 0.2 pixels^{-1} which corresponds to spatial scales of about five pixels. This extra noise at small scales comes from horizontal pixel area variations observed in Figure B.2. The noise would actually appear larger here however capacitive coupling has an effect of smoothing out small scale noise somewhat.

The power spectra fits were performed in two ways. First, the largest scales were modeled by a power law in the wave number (k). Only the isotropic power spectrum at wave numbers smaller than 0.09 was fitted. Because close to that value the white noise starts to have a small impact, the white noise contribution was fixed to 0.9% in order to allow the power spectrum to have some curvature there. In practice the procedure is as follows. A realization of the model LSS power spectrum, P_{LSS}, in created and the 18^o line feature is added to it. Then the isotropic power spectrum of the model realization is computed and fit to the isotropic power spectrum of the data. The large scale noise power is parametrized as such:

$$P_{LSS} = p_{norm} \times \left(1. + \left(\frac{100}{k}\right)^{\nu}\right) \ , \tag{B.2}$$

where p_{norm} is the normalization fixed at a wave number of 100 and has a value of 4.96×10^{-6}. The power law slope ν has a fitted value of 2.16. This fit is shown as the blue line in the left panel of Figure B.7. Green and red lines show the effective vertical and horizontal one dimensional power spectra using the best fit models but were not themselves fit.

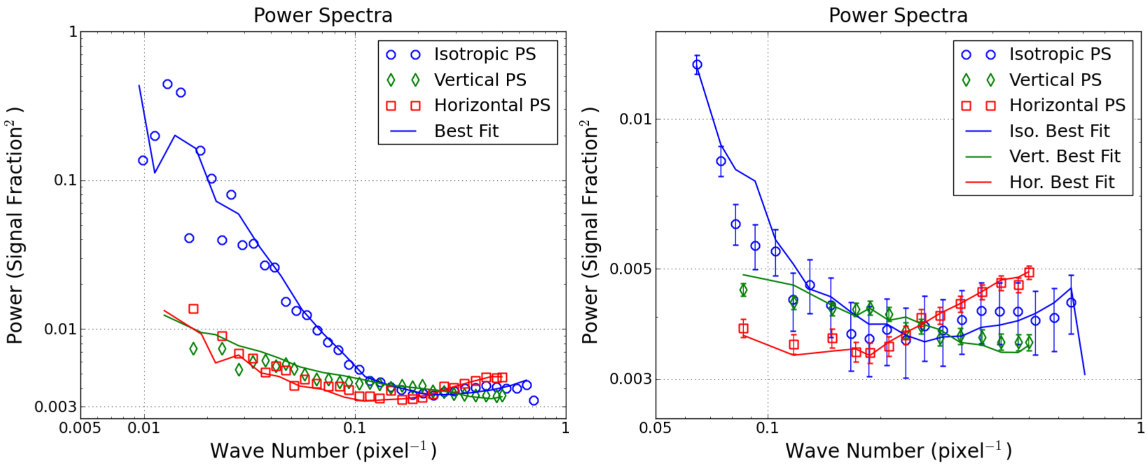


Figure B.7: Power spectrum of the entire flat detector image on the left and of small scales (64 × 64 pixel squares) on the right. The LSS was fit without uncertainty estimates. The SSS uncertainty estimates come from the use of 25 smaller square regions.

The Small Scale Structure (SSS) power was fitted assuming the best fit model of the large scale structure computed above. The small scales were assumed to be a combination of an irreducible random noise floor, random pixel area variations in the horizontal direction and a small reducible random noise left over due to a finite number of frames being averaged. The area variation model assumes that the left and right edges of each pixel are displaced from the nominal values by an amount drawn from a Gaussian distribution with a mean of zero and width σ equal to the area variation noise contribution divided by $\sqrt{2}$. The reducible random noise contribution was calculated using the difference between the small scale noise in individual frames as well as in the averaged image and was therefore not fit for.

Model realization of the LSS obtained above was combined with the 18^o line feature model, the irreducible and reducible random maps and the area variation model. Finally, capacitive coupling is applied and the power spectra are computed. The standard deviations of the irreducible random noise and of the pixel area variation model are then adjusted until they fit the true small scale power spectra. Unlike the LSS above, isotropic, vertical and horizontal power spectra were all fit simultaneously since all three constrained the model in different ways. For example, the horizontal power spectrum contains much of the information on the horizontal pixel size variations. The results of the fit give a irreducible random noise of 0.73% and horizontal area variation noise of 0.68% BEFORE the application of the known capacitive coupling value of 2.2%. That capacitive coupling reduces the small scale noise by about 8%.

B.4 H2RG-103 Results

In this section we check the fidelity of the model by comparing the outputs of our various tools run on realizations of the model and the data itself. Table B.1 summarizes the model parameters.

18^o Line Feature	Amplitude Distribution* σ	0.2%
	Angle Distribution σ	3^o
Large Scale Structure	Power Law Slope	2.16
	1/2 Power normalization at k=100	4.96×10^{-6}
Small Scale Structure**	Irreducible Random Background	0.73%
	Hoizontal Pixel Area Variation	0.68%

Table B.1: The H2RG-103 detector noise model parameters derived in section B.3.
* Amplitude was constrained to negative random draws from a Gaussian distribution only.
** Before applying capacitive coupling.

The full comparison required a simulation of an entire image set. Once an independent realization of the background irreducible sensitivity structure of the same size as the flat detector image was created, one hundred “exposures” were generated by adding random shot noise to the irreducible background. Each image then had capacitive coupling applied. These one hundred images were then combined to create an average frame that could be compared to data (see right panel of Figure B.8). Note that for the angular noise spectrum, fifteen additional independent realizations of the background were created in an identical manner.

B.4.1 Image

Figure B.8 shows the real detector flat image section on the left and a realization of the model on the right. The large scale structure is not as similar as we would like. This is partially due to the fact that each random realization will be different. But in addition the detector image shows more structure in the horizontal direction which is consistent with the higher horizontal power (red squares) in the left panel of Figure B.7. Nevertheless, I believe that at small scales, where we are most interested in the detector response, the model does much better.

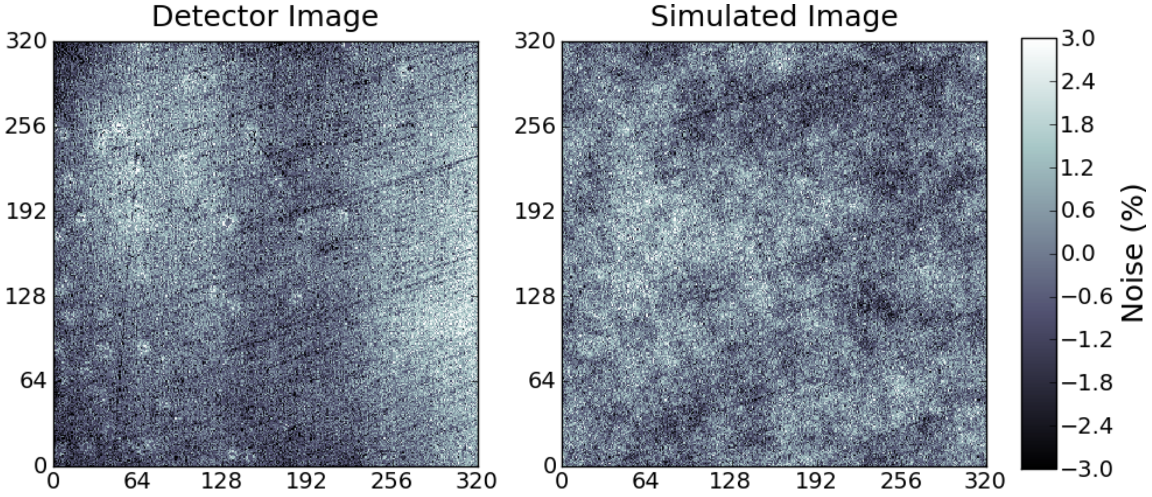


Figure B.8: The actual (flat) detector image on the left. Simulated image on the right. The images were divided by their mean values, one was subtracted from them and they were multiplied by 100% to produce noise maps.

B.4.2 Correlation Function

The correlation function analysis was run on a realization of the simulated image in order to compare it to the analysis performed on real data and shown in Figure B.2. Figure B.9 shows the correlation functions computed with high pass filtering at lengths larger than 16 pixels while Figure B.10 shows the correlation functions computed with high pass filtering at lengths larger than 64 pixels. Immediately one can see that in both figures, the simulated anti-correlation of adjacent pixels (right panels of figures) is very similar to the anti-correlation in the data (left panels of both figures). In addition, the 18^o line feature is also clearly visible though not quite identical to the real data. The simulated line tends to spread out farther away from the center while the real one remains fairly narrow. This suggests that the assumption of a distribution of angles centered at 18^o with σ=3^o was likely incorrect. Instead, the lines themselves may be "thicker" than one pixel leading to the width observed in the angular noise spectrum of Figure B.4.

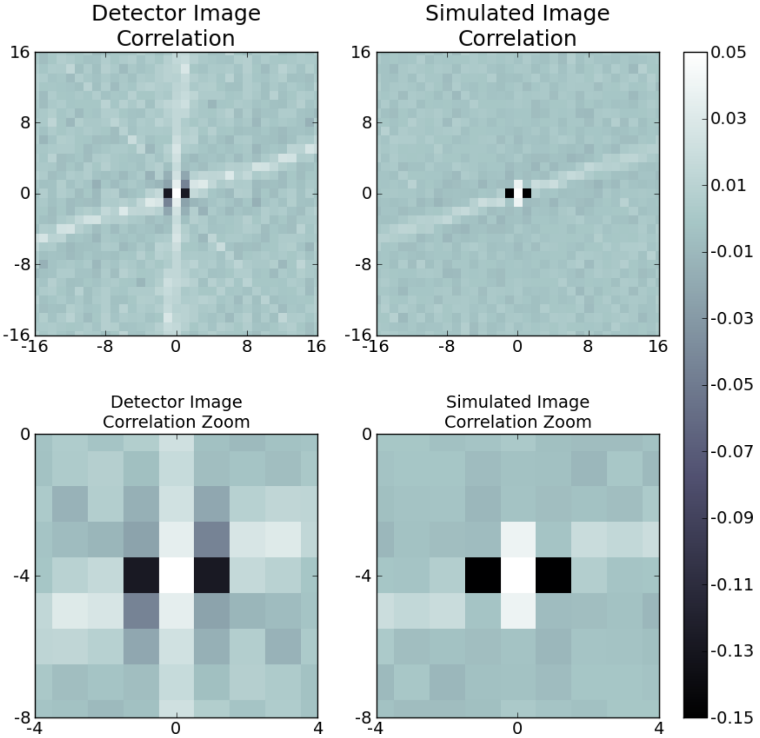


Figure B.9: Correlation functions of the data (left panels) and simulated image (right panels). Lower panels are the zoomed in versions of the upper ones. These correlation functions extend 16 pixels in each directions and were obtained with high pass filtering of scales larger than 16 pixels.

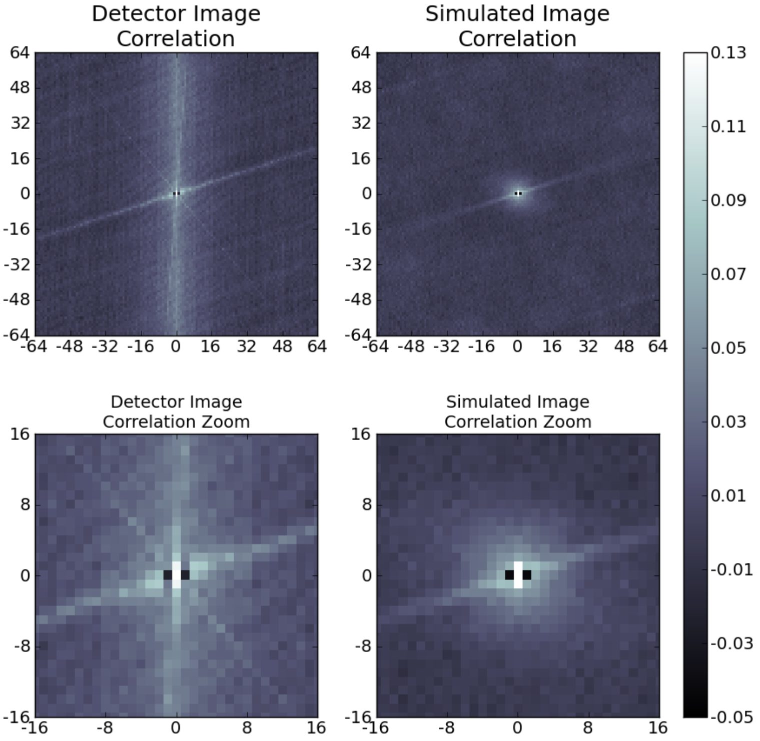


Figure B.10: Correlation functions of the data (left panels) and simulated image (right panels). Lower panels are the zoomed in versions of the upper ones. These correlation functions extend 64 pixels in each directions and were obtained with high pass filtering of scales larger than 64 pixels.

B.4.3 Angular Noise Spectrum

Figure B.11 shows the angular noise spectrum of the data (as a blue line) and the region spanned by 16 independent realizations of our noise model. The wide dip around 0^o is the result of horizontal pixel size variations. Vertical pixel size variation, when present, would cause similar dips at 90^o and -90^o. Again the 18^o line feature is clearly visible in the simulations. The simulations qualitatively resemble data though the match is certainly not perfect. Besides missing some linear features, our model does not reproduce the behavior of the data at 50^o well. This may be related to the imperfection of our modeling of the 18^o line feature or it could be unrelated.

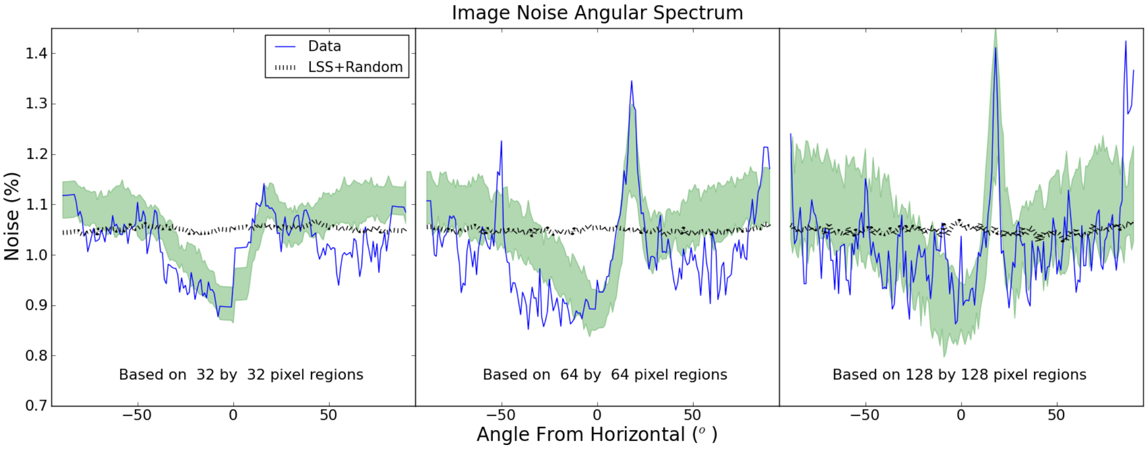


Figure B.11: Angular noise spectrum. See the caption of Figure B.4 for the description of the 3 panels. The spectrum of the real image is shown as a blue line. Dotted black line shows the expectation from random small scale noise. The green band shows the 1σ region where the 16 independent realizations of our noise model lie.

B.4.4 Multiple Exposures Averaging

By averaging multiple frames of the simulated images normally as well as after partial vertical and horizontal averaging we see in the left panel of Figure B.12 that our model (green circles, dashed and dotted lines) reproduces the data (blue $\times$ symbols, dashed and dotted lines) quite well. A random model with capacitive coupling is once again shown in black.

In order to look at these results in more detail, the right panel of Figure B.12 shows the ratios of the vertically partially-averaged noise reductions to normal noise reductions as dashed lines and horizontally partially averaged noise reductions to normal noise reductions as dotted lines. The data (blue) and model (green) agree quite well. The horizontally partially averaged ratios are lower than the vertically partially averaged ratios because of the horizontal anti-correlation introduced by the pixel size variations. Note that the shot noise model ratios (black) are slightly higher than 0.5. This is caused by capacitive coupling which introduces small positive correlation between adjacent pixels making the noise reduction effect of averaging four pixels in a column or row (that is, partial averaging) smaller than a factor of two.

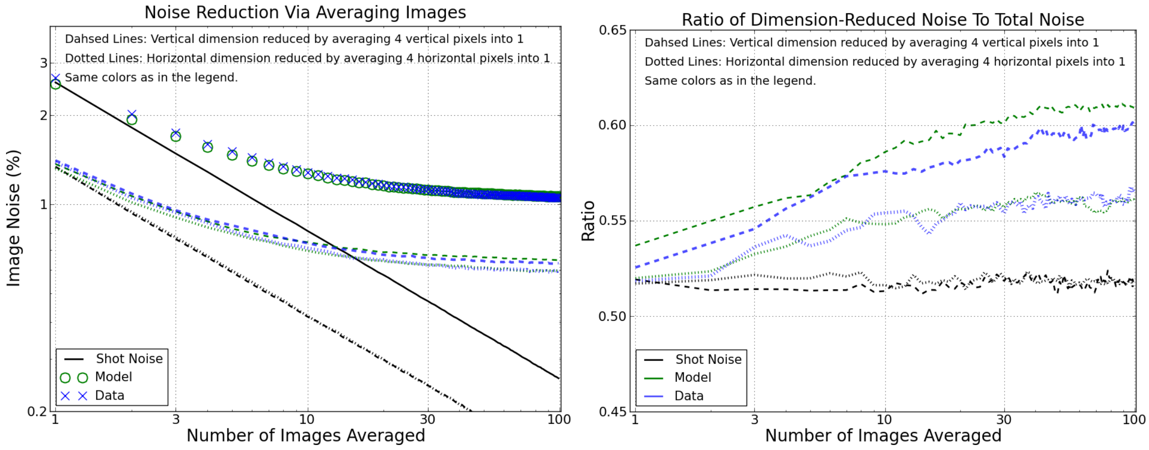


Figure B.12: Left panel: Noise reduction via averaging multiple exposures. In addition to the overall noise reduction (blue ×), partially-averaging individual exposures in vertical (dashed) and horizontal (dotted) directions is also shown. Green circles and lines shows the results of performing the same analysis on simulated exposures. Predictions based on shot noise are also shown as black solid, dashed and dotted lines. Right panel: Ratios of the partially averaged data and simulations relative to the normal noise reduction case.

B.5 Model limitations

Our model of the detector sensitivity variations is not complete. Some of the notable features that are missing are discussed below.

B.5.1 Eighth Column Correlation

As indicated in Figure B.3, every eighth column appears to be correlated within itself as well as with the zeroth column. That figure is based on the full detector correlation function which may bias our results. However we confirmed that this effect is also visible in the flat detector image although at a lower significance indicating that it is not a strong effect. We suspect this effect may be caused by an offset in the placement of the photo-lithography mask and therefore would be classified as a column width variations and should not be flat fielded away. However this is not certain.

B.5.2 Additional Line Features

Both the correlation functions in Figures B.9 and B.10 as well as the angular noise spectra in Figure B.11 show that there are additional line features than just the 18^o one. Note that the eighth column correlation discussed above is one of them though we suspect that unlike it, the other features are not area variations but rater sensitivity variations. Since we do not know exactly how to describe them and they are significantly weaker than the 18^o line feature they were left out of our model.

B.5.3 Capacitive Coupling

Capacitive coupling of pixels (Moore et al. 2004; Brown et al. 2006; Barron et al. 2007) introduces small positive correlation between adjacent pixels because it transfers a small fraction of the collected electrons from the brighter to a dimmer pixel. It therefore has an effect of smoothing out structure. We are not able to fit for it with the data we have available since it is degenerate with the random noise and area variations. As shown in Figure B.13 the power spectra of images without capacitive coupling (dashed lines) have an overall similar shape as the power spectra of models including capacitive coupling. They have different normalizations but an increased capacitive coupling can easily be accommodated by decreasing the random noise power. Independent measurements of capacitive coupling were therefore necessary in this work.

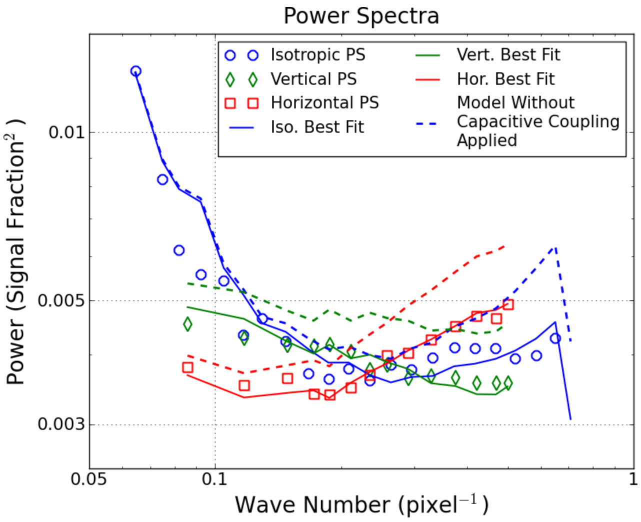


Figure B.13: Small scale power spectrum. Also shown is the model without CC applied. Note that error bars are not shown for clarity.

B.6 Summary

We have presented the results of a statistical analysis of an HgCdTe NIR detector dubbed H2RG-103 using novel tools developed for this study. We have concluded that at small scales, a few pixels, the non-uniformity of the device is dominated by irreducible pixel sensitivity variations and horizontal (read direction) pixel area variations. Both of these effects are suppressed by capacitive coupling hence their magnitude can only be obtained using an independent measurement of the capacitive coupling strength. That being said, the *presence* of horizontal pixel size variations is clear. On larger scales, about 50 pixels, the noise power is relatively isotropic with a $1/f^{2.16}$ frequency dependence. We also detect and model *linear* features in the device most likely due to the manufacturing process. As an aside we note that a brief attempt was made at applying this analysis to a different NIR detector, the H2RG-236. While many features were detected statistically our modeling was unsuccessful since their behavior did not match our hypothesized mechanisms. This device has a more complicated structure that lends itself to an analysis using tools like the Spots-O-Matic instead.

APPENDIX C

The Spots-O-Matic

Under-sampled detectors can suffer from various systematic uncertainties due to the unknown pixels response function (PRF). For point sources the problems arise due to the gradient of a star's point spread function (PSF) changing significantly over an area of a pixel. Portions of a pixel where the response is lower than average will suppress the contribution from light falling onto them and vice versa since regular flat fielding will only account for the average of the PRF. For sources like under-sampled galaxies, the inhomogeneity of PRF will increase the measurement noise at best and imprint additional structure in the galaxy image at worst. In addition, the PRF can be *effectively* altered by the underlying electronics and observational conditions. This emerged clearly in our measurements and will require far more study than what is presented here.

C.1 The Apparatus

The principle of the Spots-O-Matic (Biesiadzinski et al. 2010) is the same as the Spot-O-Matic (Barron et al. 2007), the single spot projector. A spot, substantially smaller that a detector pixel, is scanned along the vertical and horizontal dimensions. The charge integrated in the pixel is recorded as a function of position of the motorized stage that moves the spot. This signal is then plotted for an *individual* pixel as a function of the position. Figure C.1 shows the schematic diagram of the Spots-O-Matic setup and Figure C.2 shows a photograph of the apparatus with the dark box removed. Unlike the Spot-O-Matic which utilized a modified microscope-based projector to generate a single spot, the Spots-O-Matic needs to create approximately 100,000 spots. A pinhole mask is therefore back-illuminated and projected onto an Mercury-Cadmium-Telluride (HgCdTe) detector H2RG-236 via a photographic lens. The detectors is mounted in a cryogenic dewar and operated at 130K. The image

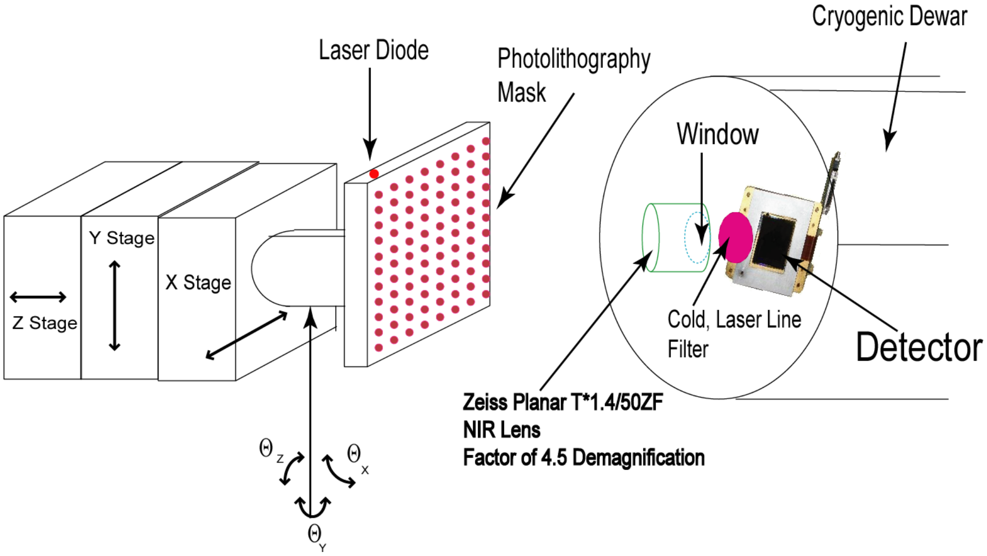


Figure C.1: Schematic diagram of the Spots-O-Matic. See text for details.

distance is fixed by having the detector connected to the front cover using an Invar mount compensated with copper to null out thermal contraction.

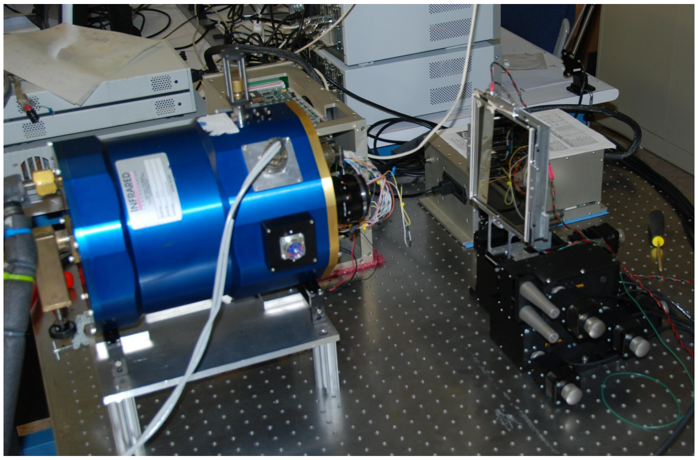


Figure C.2: A photograph of the Spots-O-Matic setup. A dark box normally covers the stage and the pinhole array (right side of picture) as well as the lens to keep out outside light. It was removed so that details would be visible.

C.1.1 Projection System

The Spots-O-Matic projects many (≈ 100,000) spots onto a detector. Each spot is small relative to the size of a pixel allowing it to probe the PRF. The spots are spaced approximately five pixels apart in order to minimize crosstalk while allowing for a speedy measurement. A 7” × 7” photolithography mask containing an array of circular apertures (the pinhole array) is illuminated by two narrow-band laser diodes centered at 1050 nm. The pinhole array is de-magnified and imaged onto the detector by a 50 mm camera lens, the Zeiss Planar T* 1,4/50 ZF IR, of low to moderate f-number. Low f-number setting reduces the diffraction-limited PSF size however it is also subject to more lens abberations hence increasing the f-number may be necessary (Bertram Hönlinger, personal communication, 2008). The resulting image is the spot array. A cold short pass filter blocks out NIR background with wavelength longer than 1100nm. Spots-O-Matic scans require that the spots be as small as possible, that is, the pinholes must be in focus. This is achieved via rough focusing, roll correction, fine focusing and additional roll correction. The design goal was to achieve spots with full width at half maximum (FWHM) of 7 μm. It is likely that the goal has not been met and the actual spot size is approximately 10 μm however it is difficult to disentangle that size from the complicated pixel structure measured during the experiment.

C.1.2 Motorized Stage

The pixel PRF is characterized by scanning a spot across it vertically and horizontally. For this purpose the illuminated pinhole array was mounted on a 6 axis stepper motor stage. In addition to Spots-O-Matic scans themselves, this stage also assures that the spots are in focus, coplanar with the detector and aligned with detector rows and columns. The Thorlabs NanoMax 604 stage satisfied our requirements (1 μm step size, 4 mm focus range, 3^o angular range and micron level repeatability) and was therefore used. The focus range proved insufficient hence manual adjustments were necessary to bring the pinhole array into rough focus. The (de-magnified) vertical and horizontal steps were 2 μm in size for CDS and 3 μm in size for SUR modes.

C.1.3 Focus

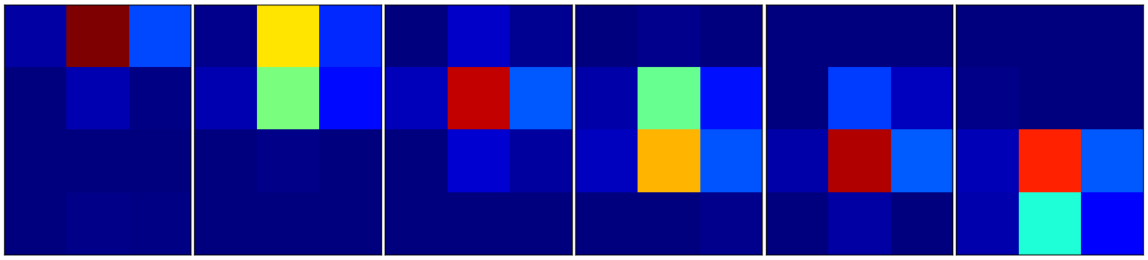


Figure C.3: A graphical example of a focus sequence. The record of a spot crossing a multi-pixel boundary is shown with the 6 top pixels forming the focus region.

The pinhole array is first roughly focused via "eyeball" detection of spots and their dependence on the focus distance. This includes adjusting both the overall object distance as well as tip and tilt of the stage. Afterward, the device roll (rotation about the longitudinal or focus axis) is derived by detecting rows of spots and computing their angle from horizontal. This roll is then removed using the roll axis of the 6 axis stage.

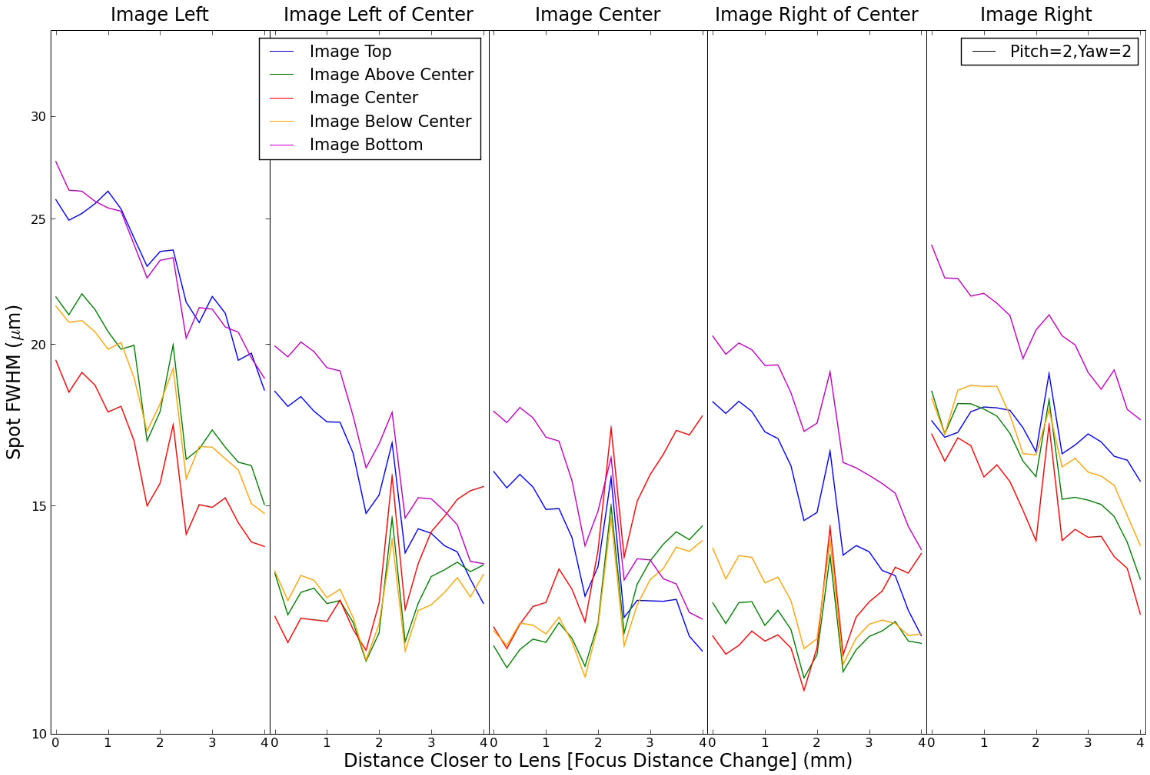


Figure C.4: Spots-O-Matic fine focusing data demonstration. The spot FWHMs are obtained by fitting an error function to virtual knife edge transitions of multiple spots at 25 locations on the detector, a 5 by 5 grid. As discussed later, the pixel PRFs can be very complicated which would degrade the spot FWHM measurement hence the values here should be treated as a spot width proxy and not the actual spot FWHM. They should nevertheless indicate the location of best focus.

Fine focusing is more complicated due to the fact that at best focus the spots are smaller than the pixels that image them. This makes it impossible to judge their size visually or by fitting the spot profiles. Instead, fine focusing is accomplished by performing virtual knife edge scans (Firester et al. 1977; Barron et al. 2007) in either horizontal or vertical direction in addition to scans in the focus direction. At a given focus positions, spots (or in this case bright pixels) are located in a first image of a series that has been scanned either vertically or horizontally. A two pixel wide and three pixel high region is summed for horizontal knife edge scans, two pixel high and three pixel wide region is used for vertical scans. These regions are defined such that the signal in the pixel furthest from the region boundary in the scan direction is higher or as high as the signal in the pixel at the region boundary in order to make sure that the spot has not yet begun to cross the virtual knife edge. Figure C.3 shows an example of a vertical scan out of one region and into another across a virtual knife edge. The six panels represent six steps of the spot in the vertical direction. A region is defined by the top 6 pixels (2 rows by 3 columns). As the spot moves out of this region into the one below, the signal in the top region decreases to essentially 0 (dark blue color). This procedure assures that spot signal is not artificially "lost" if the PSF is too large and extends beyond a single pixel.

The effect is a curve that starts high and then smoothly transitions to low. The goal is to minimize the width of this transition by moving the stage either closer to or further away from the lens. These minima will occur at different focus distances for different parts of the spot array for two reasons; the spot array box may not be co-planar with the detector and the focal surface is curved. In order to make the detector and spot array co-planar, the virtual knife edges are measured in different sections of the detector. In each section the average of multiple spot transitions is taken in order minimize statistical noise and then the different sections are compared. Figure C.4 compares 25 such sections located on a 5 column (left, left of center, center, right of center, right) by 5 row (top, above center, center, below center, bottom) grid. The 5 horizontal panels of the figure stand for the 5 horizontal image analysis regions and the 5 colors stand for the 5 vertical regions. While the image center looks to be at best focus at a relative focus position of 0 mm, the outer sections of the device (horizontally: left and right panels, vertically: blue and magenta colors) are far from focus. Through trial and error tests it became clear that the entire image plane could not be brought into focus due to its curvature. Hence, we restricted our analysis to the 1024 × 1024 pixel region in the image center, a quarter of the total number of pixels. All subsequent references to the detector are therefore limited to this region unless otherwise noted. The outer regions of the device where brought into focus separately and data was obtained however not analyzed beyond simple consistency checks due to limits in time. Nevertheless, with 2 or 3 focus positions all pixels on a device can be characterized within a few weeks.

C.2 Exposure Time and Flux

Many thousands of images are taken as part of Spots-O-Matic data acquisition in order to sample each pixel with multiple spots leading to effective dithering. Data was acquired in two sampling modes. Correlated Double Sampling (CDS) images are the difference between a read frame and a pre-read frame with the time between the two being the exposure time. Unfortunately, since our system operates without a shutter light is incoming onto the device during the pre-read meaning that the CDS image signal is lower than the actual well fill of a given pixel. All of the CDS results presented here use a 1.5 second exposure time. The illumination was not constant. The signals recorder by pixels (spot + background) ranged from 5000 ADU to 9000 ADU. The 1.5 second exposure time background is shown in Figure C.8. On the other hand, Sample-Up-the-Ramp (SUR) images consist of the difference between a 6th and 5th frame reads, 5th and 4th reads, ..., second and first reads, first and dark reads. There are therefore six images for an SUR set showing the *change* in the integrated signal between each read. For these measurements the flux was lowered by approximately a factor of 12 and each image was exposed for 4 seconds leading to a total exposure time of 24 seconds. This allowed us to see how the PRF evolved with increased integrated signal. As a check, 1.5 second CDS exposures where taken at the reduced flux. These were consistent with the first SUR images and were not used for further analysis.

C.3 Individual Pixel Processing

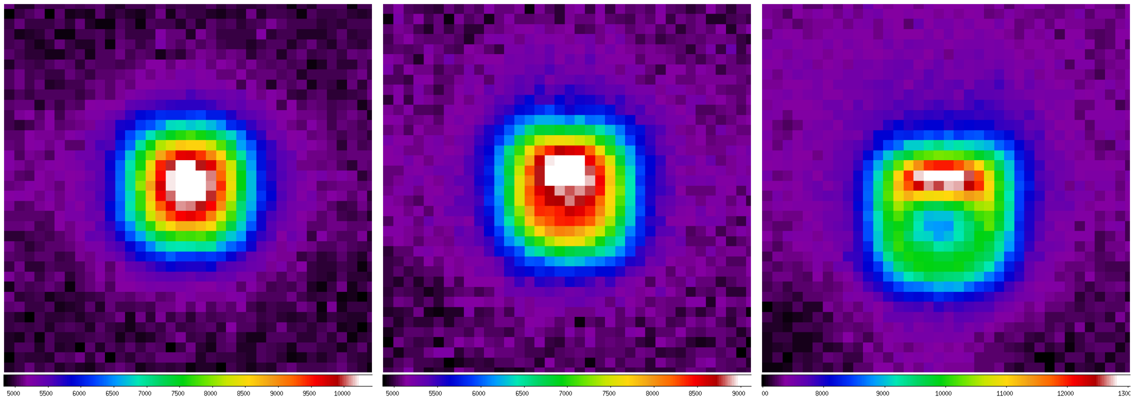


Figure C.5: Raw Spots-O-Matic measurements for three pixels. A “good” pixel is shown on the left, a “poor” pixel is shown in the center and a “bad” pixel is shown on the right. The vertical and horizontal axis are the relative locations of the motorized stage that moves the spots at (demagnified) 2 micron resolution. The color-scale is in units of ADU.

In principle, we must scan over 5 pixels horizontally and 5 pixels vertically in order to cover all pixels. This assures that each pixel is scanned by a single spot once.

However, while a single scan provides a general idea of the pixel structure (see Figure C.5) there is substantial measurement noise.

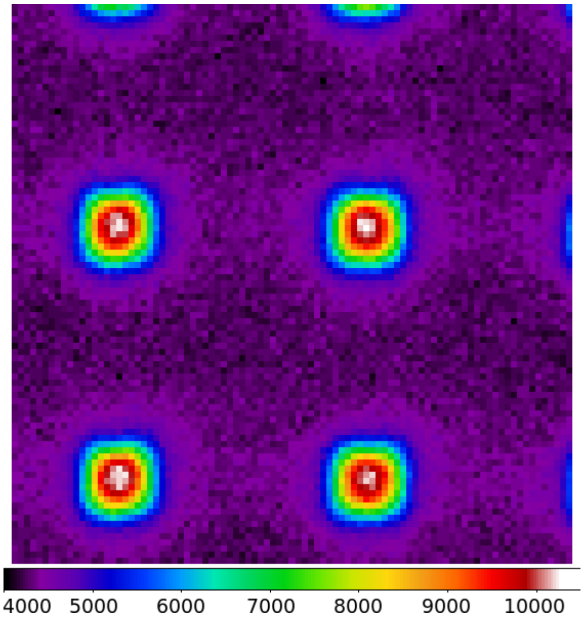


Figure C.6: Raw, single pixel Spots-O-Matic data. The muliple imgages visible are the same pixel being imaged by four different spots. Raw images like these are cut into individual spot scans and stacked together. The vertical and horizontal axis are the relative locations of the motorized stage that moves the spots at (demagnified) 2 micron resolution.

In practice, a scan therefore covers many more steps so that each pixel is sampled by several different spots. Figure C.6 shows an example of such a scan where there are four (useful) spot scans over a single pixel. A computer algorithm is used to select the individual spots from such a scan and stack them together. This also allows other scans to be combined together. The algorithm works via a rudimentary application of a Matched Filter (MF) (e.g., Herranz et al. 2002; Melin et al. 2006). The scan output for each pixel (like the one in Figure C.6) is convolved with a square of 18 μm on each side (the pixel pitch) and a Gaussian PSF model with FWHM of 7 μm. Noise was assumed to be largely *white* and therefore was not included in the MF. The convolved image was then searched for highest signal regions (pixel center as illuminated by different spots) that were at least 10 μm away from each side or any defective columns and rows due to external issues during image acquisitions. An approximately pixel sized region was masked out around each found *center* to prevent false detections. Visual inspection of data for several pixels combined with final catalog spot checks found that no false detections occurred and all available spot illuminations are recovered. The data was then up-sampled to a 1 μm resolution. Thanks to the large number of images and the lack of correlation between the stage steps and pixel boundaries we ended up effectively dithering the images. The centers of each pixel as imaged by multiple steps were used to stack the data from each Spots-O-Matic sequence and also enabled multiple sequences to be combined. Figure

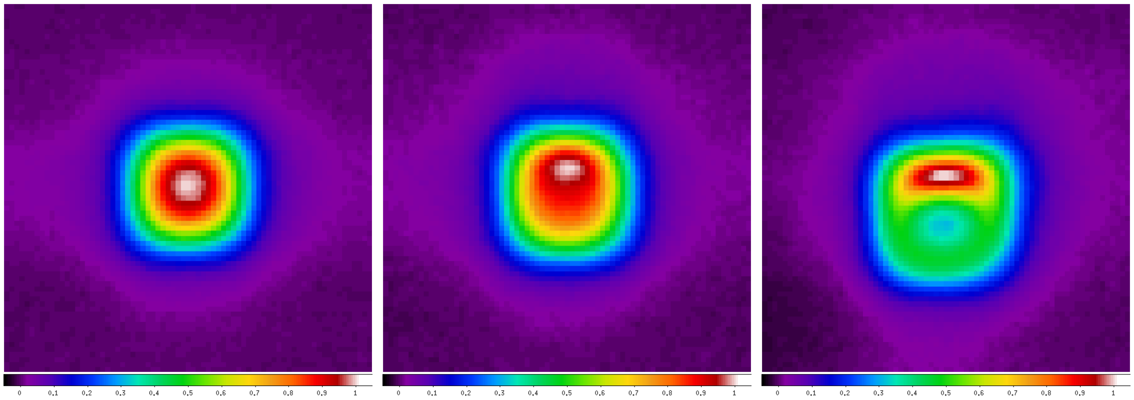


Figure C.7: Average PRF of the same pixels as shown in C.5 after stacking approximately 40 spot scans. The color-scale is normalized to the highest response region of each pixel. The vertical and horizontal axis are the relative locations of the stage at (demagnified) 1 micron resolution.

C.7 shows the images for the same 3 pixels as Figure C.5 but each one created by averaging over 40 spot scans.

C.4 CDS Results

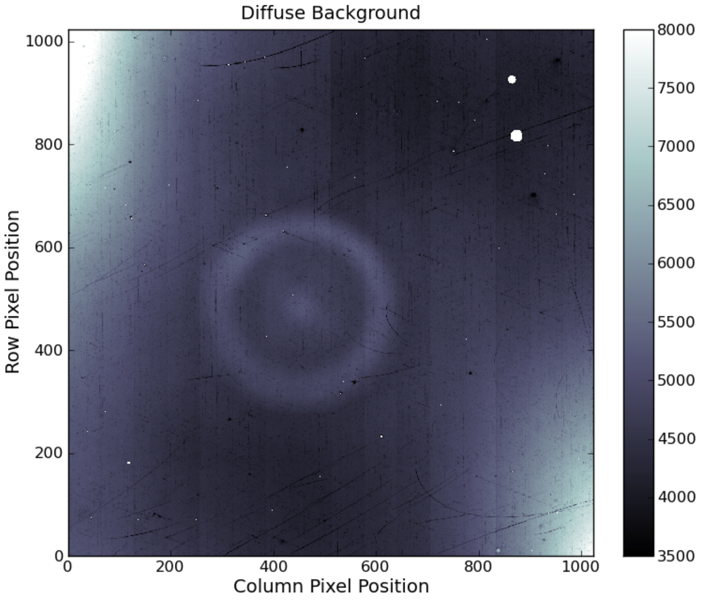


Figure C.8: Diffuse background illumination in Spots-O-Matic CDS measurements. It was obtained by looking at the outer edges of each PRF, far away from the active area of the given pixel. The mean of these edges was plotted for each pixel.

As shown in the previous section, the pixels in H2RG-236 can exhibit a great deal of structure. The overall detector sub-pixel structure can be characterized via different means. The first thing to consider is the background illumination. Light leaves the pinhole array at various angles due to the shallow depth of our illumination box. Combined with diffracted light, leakages and reflections, the background signal is significant as shown in Figure C.8. There are several obvious features. The top left and bottom right corners are significantly brighter. These are the locations where the laser diodes are mounted in the light box hence they are illuminated more. Individual readout channels also vary slightly in their intensities. This is due to small variations in channel gains and are not an issues for us. The rings visible in the center are reflections of the ring around the lens barrel as well as the lens itself. In addition, white pixels and small white regions are due to masking of bad pixels. We refer to this image as the *diffuse background* because it is due to light that is uniform on small scales (dozens of pixels).

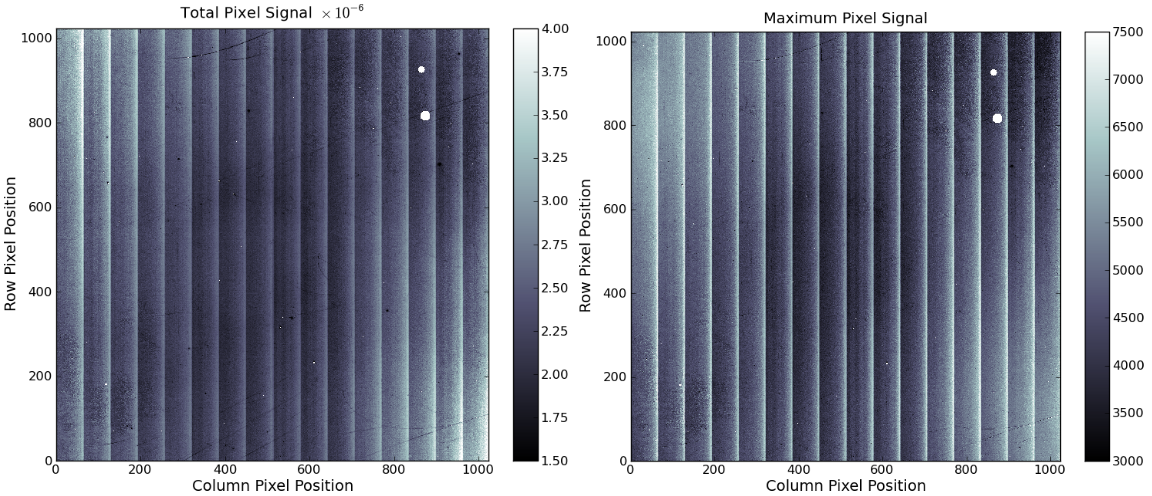


Figure C.9: Total and maximum spot signals. Left panel: the total PRF in ADU summed for each pixel. Right panel: the maximum of the PRF of each pixel. Both are calculated after diffuse background subtraction.

Figure C.9 shows the total signal for each pixel in the left panel and the maximum signal for each pixel in the right panel. Note that the diffused background has been subtracted prior to constructing these two maps. The top left and bottom right corners register higher signal due to the larger illumination in those regions. However, the rest of the structure is far more complicated. By far, the most obvious feature is associated with the readout channels. The right side of each channel has a larger total as well as maximum signal than the left side. Additional structure is present (bottom left, top right corners and bands around rows 300 and 600) however it requires different map making approaches and is therefore discussed in §C.4.2.

C.4.1 Readout Channel Structure

We shall first discuss the channel readout structure in more detail. As indicated in Figure C.9 this structure appears in each of the 16 readout channels shown. In Figure C.10 we plot 8 sample columns in two different readout channels, each in a different row. The upper panels show several of the 64 pixels spanning the readout channel at row 512, essentially the center of the device. Specifically, it shows pixels 0, 9, 18, 27, 36, 45, 54 and 63 of the 9^{th} readout channel shown (17^{th} channel in the full device readout). The bottom panels show the same order of pixels relative to the first shown readout channel (9^{th} channel of the full device readout) in row 10 of the map. Several features should be pointed out. The pixels in the first two columns (only the first one is shown at the very far left) of a readout channel are fairly symmetric but the following ones get considerably more distorted until after past the midway point where they start becoming more symmetric again. The rightmost pixels look as expected, a square pixel convolved with a round PSF. Furthermore, the lower panels show larger degrees of distortion. This is correlated with position in the device. The lower left and upper right sections of the device show more distortions and this will be discussed shortly.

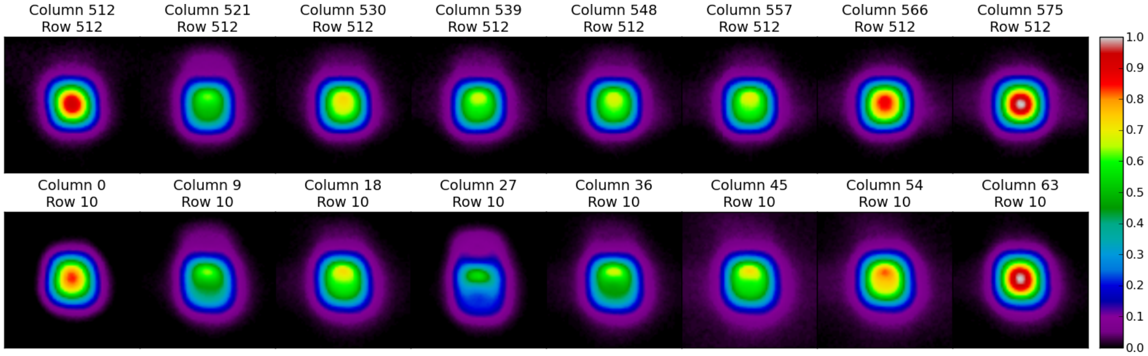


Figure C.10: Sample pixel PRFs from two rows of two different readout channels. See text for details.

C.4.2 Gross Detector Properties

Further analysis of the large scale detector features employed a technique meant to differentiate effective pixel extents in the vertical and horizontal direction. For this reason, the response of each pixel, like those shown in Figures C.7 and C.10, are normalized such that the total pixel signal is unity. Essentially, each pixel PRF is divided by the total value in the left panel of Figure C.9. Then, the normalized

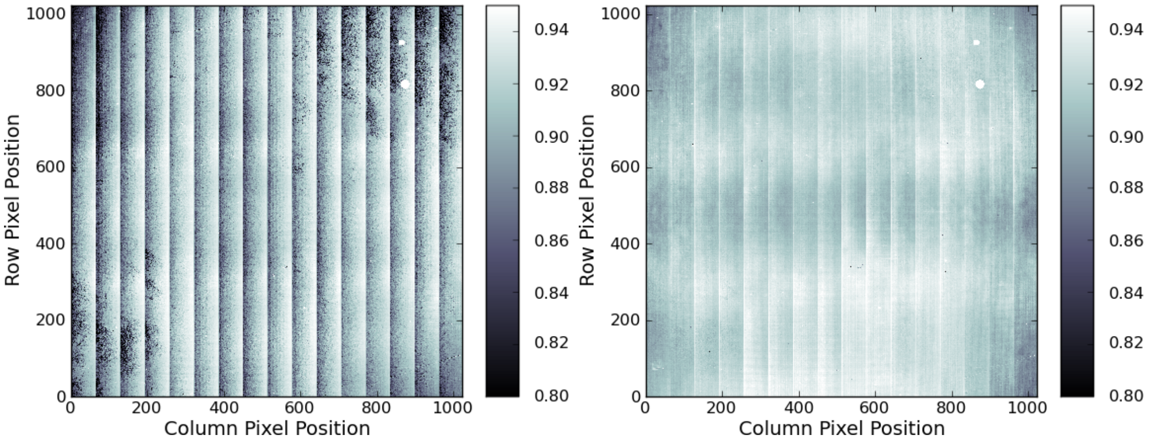


Figure C.11: Vertical (left panel) and horizontal (right panel) 30 μm wide PRF integrals. See text for details.

response of each pixel in a strip 30 μm high or wide and centered on pixel center is summed. The results are shown in the left panel of Figure C.11 for the 30 μm high stripe sum and in the right panel of that figure for 30 μm wide stripe sum. If the image of the pixel fit within the 30 μm band its response would be unity. Lower values indicated that more of the pixel effective area extends beyond that band. Figure C.11 shows that the 30 μm wide stripe sums (right panel) have far less structure than the 30 μm high stripe sums (left panel). This is consistent with what is seen in Figure C.10 since the *bad* pixels tend to extend vertically and not horizontally. There is some large scale structure visible in the right panel, specifically three darker bands extending between rows 50 to 200, 350 to 600 and 700 to 900. This indicates that the effective pixel width in those bands is slightly larger than at other locations.

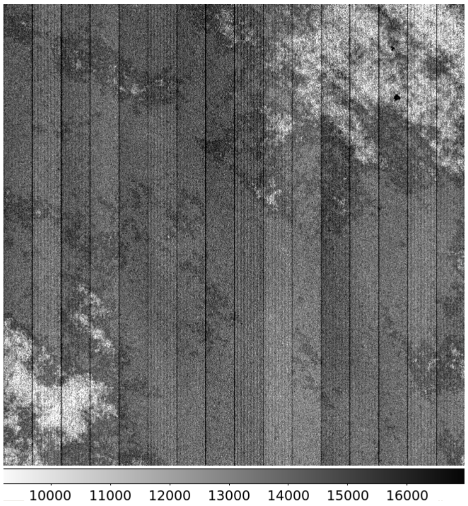


Figure C.12: H2RG-236 bias voltage snapshot in ADU.

The PRF summed in the vertical direction (left panel of Figure C.11) shows far more structure with a larger dynamic range. The unique patterns in the bottom left and top right corners indicating a *taller* effective pixel area there were immediately reminiscent of the detector bias voltage picture shown in Figure C.12. Meanwhile, the top left and bottom right corners match the location of increased illumination due to laser diode positioning as seen in Figure C.8 which suggests that the vertical extent of the pixel may also be a function of flux or total integrated signal (§C.5 will discuss this in further detail).

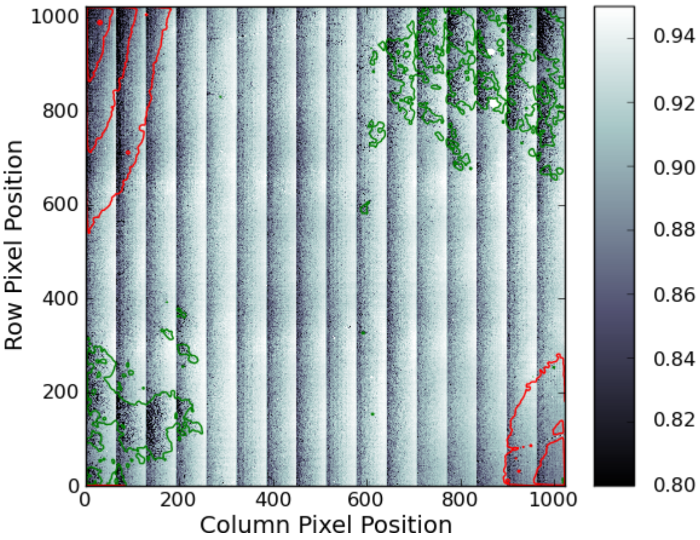


Figure C.13: Vertical 30 μm wide PRF integral overlaid with bias voltage and diffuse background illumination contours. Bias voltage contours of 12,500 ADU are shown in green and diffuse background contours of 6000, 7000 and 8000 ADU are shown in red.

To illustrate this correlation we re-plot the left panel of Figure C.11 with green lines marking the detector bias voltage of 12,500 ADU and stacks of red lines indicating diffuse background illuminations of 6000, 7000 and 8000 ADU (the 8000 ADU level is seen only in the top left corner) in Figure C.13. The agreement is quite good. To see more detail we zoom in at the bottom left corner as shown in Figure C.14. Note that the rightmost edges of each readout channel does not show the bias voltage pattern but it is very apparent to the left. The bias voltage levels are drawn at 10,000 ADU (blue lines), 11,500 ADU (green lines) and 13,000 ADU (red lines). The agreement is very good in the left halfs of each channel.

For completeness, the bias voltage and illumination contours from Figure C.13 are also re-drawn on the maps from right panel of Figure C.11, the 30 μm wide stripe sums map, and shown in Figure C.15. There appears to be no correlation further indicating that the effective pixel width is not strongly affected by the detector bias voltage, readout sequence or integrated signal.

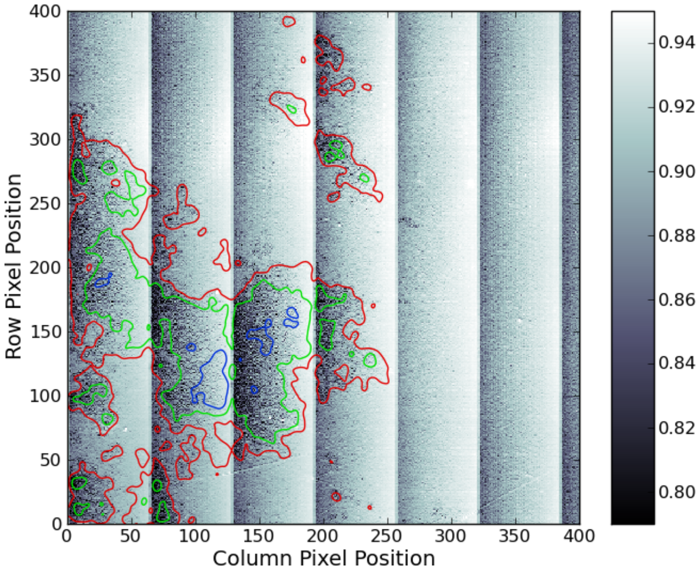


Figure C.14: Zoomed in vertical 30 μm wide PRF integral overlaid with bias voltage contours. The bottom left section of the analysis region is shown. Bias voltage contours of 10,000 ADU 11,500 ADU and 13,000 ADU are shown in blue, green and red, respectively.

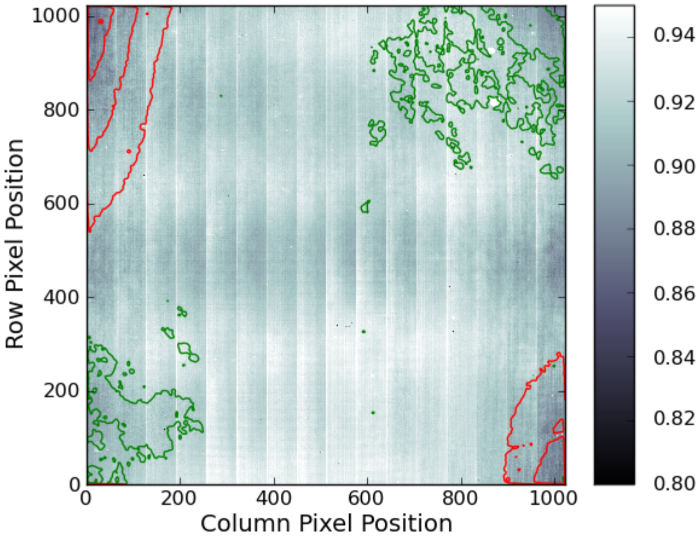


Figure C.15: Horizontal 30 μm wide PRF integral overlaid with bias voltage and diffuse background illumination contours. Bias voltage contours of 12,500 ADU are shown in green and diffuse background contours of 6000, 7000 and 8000 ADU are shown in red.

C.4.3 Pixel Sizes

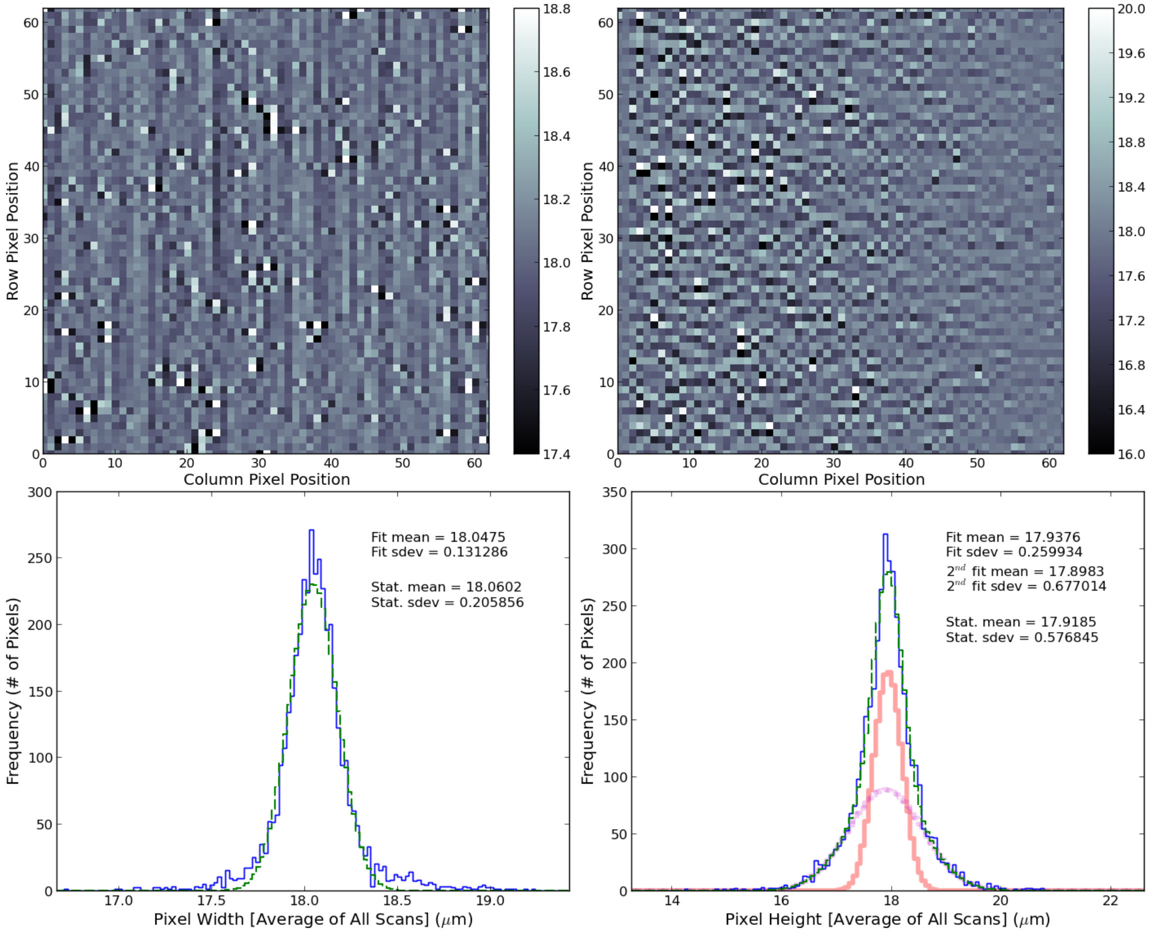


Figure C.16: Pixel width and heights, maps and distributions. A 64 × 64 pixel region in the 9^{th} readout channel (17^{th} of the full detector) was used for this study. Pixel widths are shown on the left while pixel heights are shown on the right. The map color scales are in μm. Bottom panels show histograms including Gaussian (dual Gaussian for heights) fits to the distributions.

Detailed scans of all pixels allow us to investigate pixel size variations. However, considering the poorly defined effective area of pixels discovered by our measurements it is non-trivial to define a pixel size. Instead, I used horizontal and vertical *separations* between pixel centers as defined by the MF as a proxy for pixel width and height, respectively. I selected a small region of the device, very near its center, to study in more detail. This is a 64 × 64 pixel region that spans the 9^{th} readout channel. My results are shown in Figure C.16. Pixel width map and histogram are on the left and the pixel height map and histogram are on the right. Both measurements match the manufacturers stated pixel pitch of 18μm quite well. The width map shows some signs of column width variations however this may be an artifact of our scanning strategy (see §C.7). The pixel height map shows more variation on the left side of the channel. This may be due to the fact that the MF centers are less well defined in regions where the PRF is badly deformed. It should be noted that while centroid-defined centers can be a few microns away from MF centers for such pixels, the separation between adjacent pixels is largely the same regardless of the center used.

Looking at individual pixels one can see correlations between adjacent pixels. Negative correlations in the horizontal direction for the width map and in the vertical direction of the height map are expected. That is, assuming that the detector size is accurately known, if a distance between the first and second pixels in a column appears large, the distance between the second and third pixels is likely to be small. This may be a real feature of the detector but in our case is likely to be induced by ill – defined centers instead of real pixel separations. If the height of the second pixel is shifted up, the separation with the pixel below will become larger and the separation with the pixel above will become smaller. Of more interest are the cases where pixels are correlated vertically in the width map and horizontally in the height map. These tend to be positive correlations and are likely to reveal the real structure of the device. Ultimately, it is difficult to characterize the effective area of pixels with complicated PRFs present.

C.4.4 Summed Response

While the exact shape of the PRF is very important for undersampled telescopes where source shapes must be recovered (whether for PSF fitting or weak lensing measurements), simpler applications like photon counting care only about whether the photons are recovered. For this reason I constructed a summed response map for the same region as in Figure C.16. This map combines the PRF of individual pixels by summing them up with centers offset horizontally and vertically by pixel separations derived in §C.4.3. It can tell us that, though displaced, photons from a source are nevertheless seen somewhere on the device. The resulting summed response is shown in Figure C.17 at a resolution of 1μm. The PRFs used for this map were not normalized. Assuming that the illumination uniformity is good for such a small area, a reasonable assumption, we can see that detector response is compromised since varying numbers of photons will be lost depending on the PSF position. Since this map is constructed using derived pixel separations, it can suffer from systematic

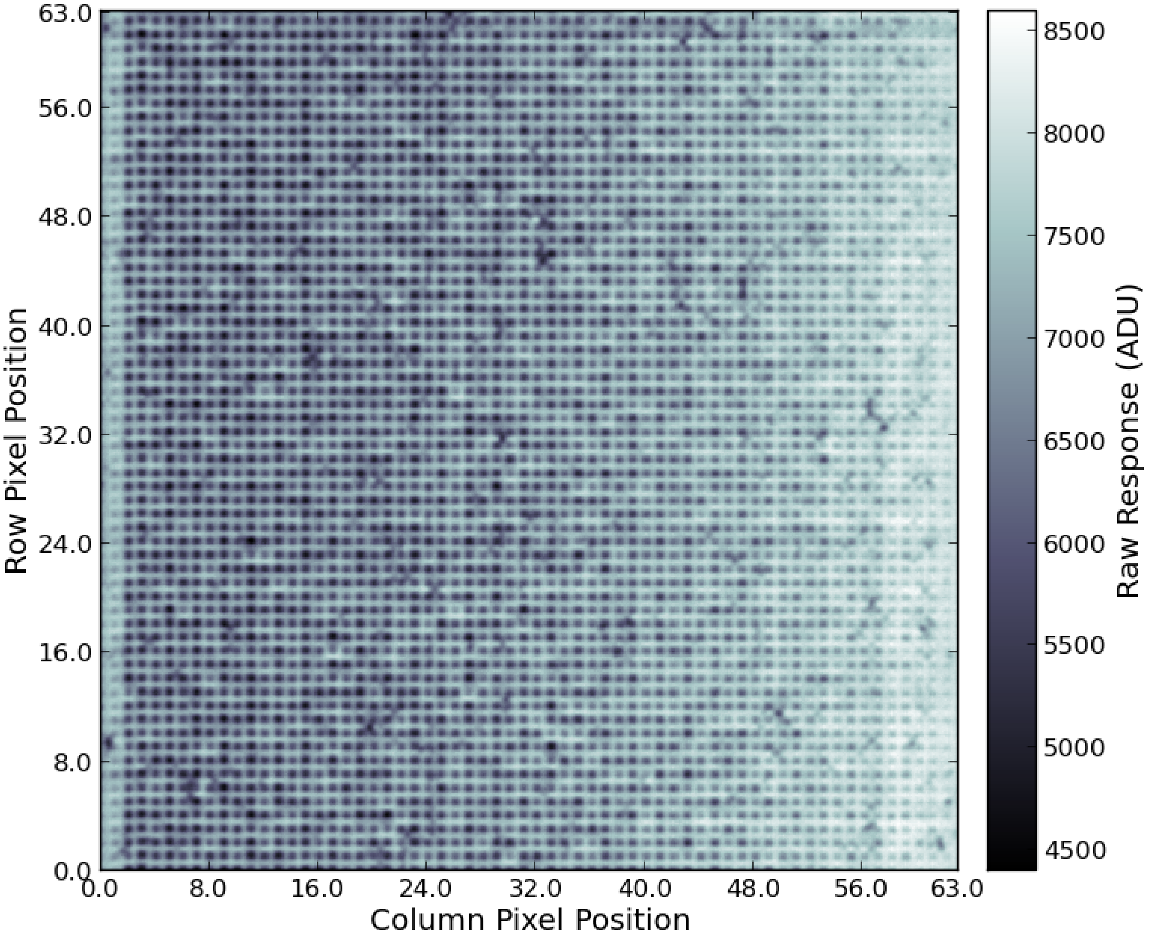


Figure C.17: Summed detector response map constructed using multiple spot scans performed by the Spots-O-Matic. The map resolution here is 1μm.

errors if these separations are not accurate. In fact, I found that rounding the center positions to the nearest micron did introduce unrealistic features to the map. The final map was constructed with pixel center positions rounded to the nearest 0.125μm to eliminate this systematic. To make sure that the procedure works correctly, in Figure C.18 I compare several small patches (left panels) cut out from the map in Figure C.17 to these patches measured directly using single spot scans in two separate measurements (center and right panels, respectively). Those two measurements do not rely on pixel separation measurements. This figure shows that structure in the constructed map is the same as in the direct measurements though it is far more precise. Interestingly, the bottom patch from all three sets shows a large defect that is not aligned with any pixel position. Finding such defects was one of the main reasons for the construction of the Spots-O-Matic and it is good to know that it can in fact do that.

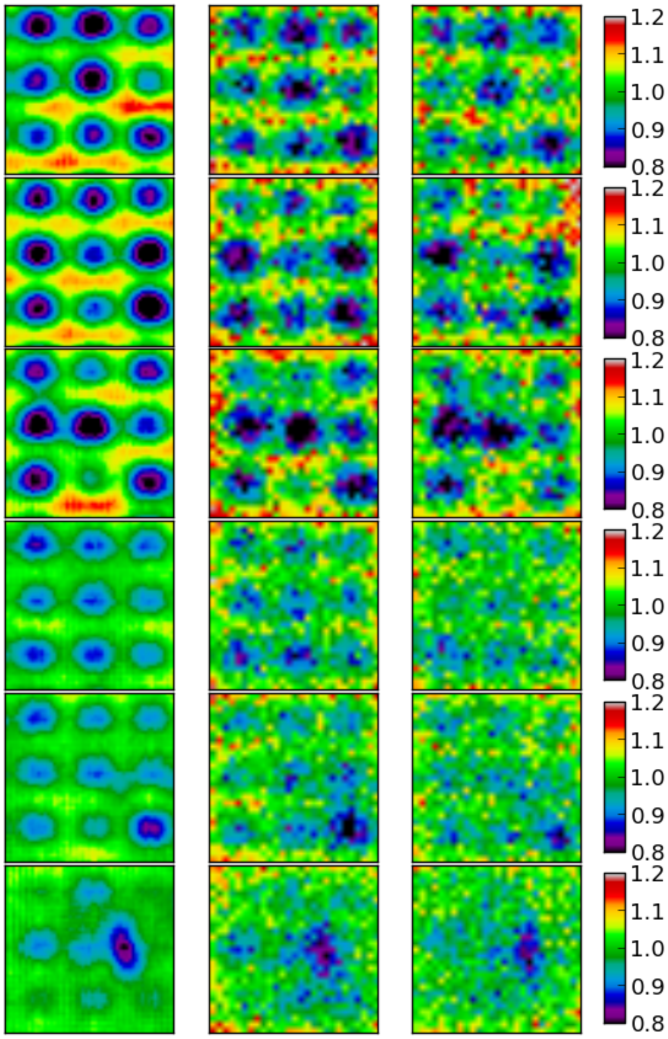


Figure C.18: Summed detector response for six small regions comparing the constructed response map to simple response maps. Shown are 54μm × 54μm regions (about 3 × 3 pixels) centered on pixels in row 514, columns 7, 10, 30, 52, 54, 61 of the ninth readout channel (17^{th} channel of the full device). The left panels show regions cut out from the constructed summed response in Figure C.17. The center and right panels come from two separate *single spot* scans obtained from summed data without any assumptions on pixel separations. All patches have had the diffuse background subtracted and are normalized to their respective median values.

C.5 SUR Results

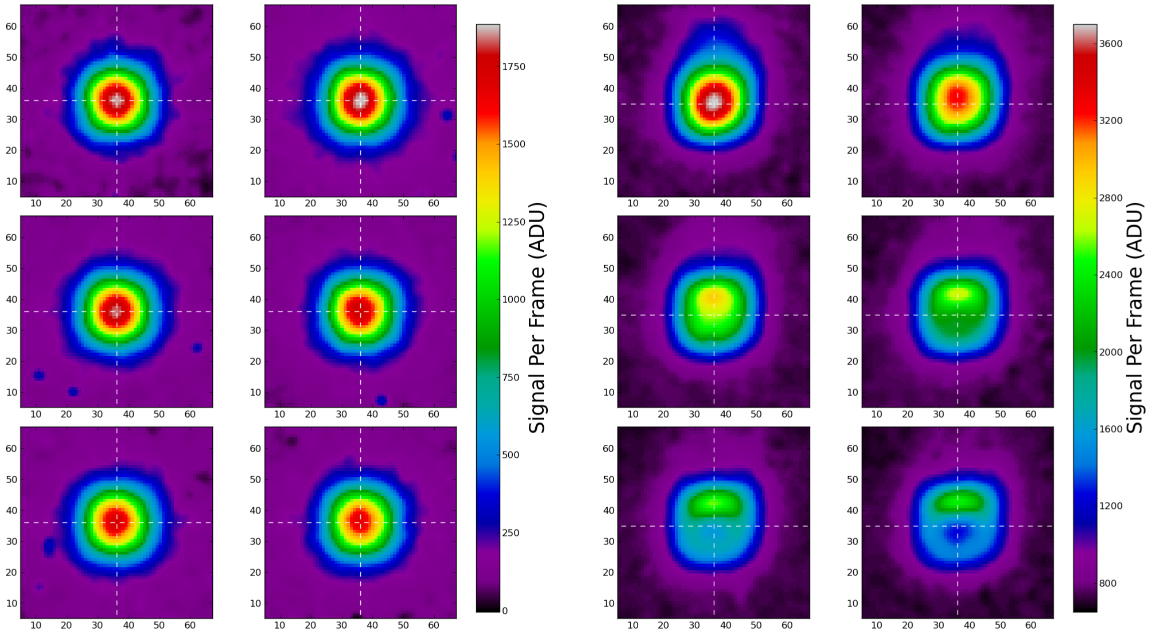


Figure C.19: SUR differential response of a good and a bad pixel in 6 frames. These due not include diffuse background subtraction hence there is non-zero signal at the edges of the PRFs. The good pixel (left) is the same one as the left panel of Figure C.7 and the bad pixel (right) is the same one as the rightmost panel of Figure C.7. The right pixel is in a region closer to the diode hence it sees a larger amount of background illumination as well as spot flux. Note that the color scales are different.

Figure C.13 indicated that the pixel shape distortion is correlated either with photon flux or total integrated signal level. To distinguish the two possibilities we undertook low flux SUR measurements with exposure times such that they would achieve roughly the same total signal level as our CDS measurements. The match is not exact since the CDS measurements include a large amount of charge collected during the pre-read which is not included in the images. Figure C.19 shows 6 differential SUR images of a “good” pixel and a “bad” pixel. It is obvious that while the shape and response of the good pixel does not change much during signal integration, the PRF shape of the bad one changes significantly and its sensitivity decreases even though the flux is approximately 12 times lower that in our CDS measurements. This explains the correlation of the PRF deformation with the background illumination since pixels in more highly illuminated regions will be altered more than ones in less illuminated regions.

It should be noted that the *total integrated* response does not look quite as dramatic as the differential response in the right panel of Figure C.19. This total response for the same pixel is shown in Figure C.20. At no point does the total signal *decrease* relative to the previous read even though the differential response decreases. In a way, this is reminiscent of the integrated signal non-linearity (see Chapter 6, §6.3.2) where the response becomes more non-linear as signal is integrated. However, integrated signal non-linearity as we understand it would be apparent in flat field illumination images which means that the bias voltage pattern would be seen in the diffuse background (Fig C.8) and in quantum efficiency measurements in Chapter 6. However this patterns does not appear there.

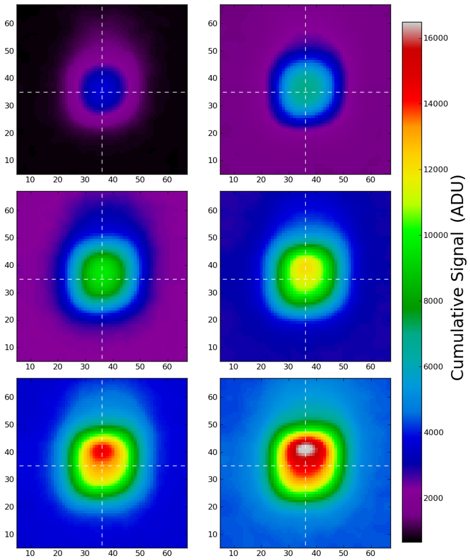


Figure C.20: SUR cumulative response of a bad pixel in 6 frames. Same pixel as the right panel of Figure C.19 is shown.

C.6 Possible Causes

Large scale variation in the right panel of Figure C.11 are likely to be actual sensitivity or area variations. They may be caused by mechanical stress experienced by the detector due to its mounting, thermal cycling or bulk sensitivity variations. It is difficult to disentangle area size from absolute sensitivity variations since the pixels in this detector behave differently under spot illumination than under flat field illumination and our spots did not have uniform brightness across the entire array. Further work is necessary to fully address these variations.

The correlations of pixel shape deformations with total integrated signal, bias voltage and readout channels were unexpected and pose significant problems to observatories using devices such as this those. It is of great interest to understand the cause of this behavior.

C.6.1 Possible Charge Transfer

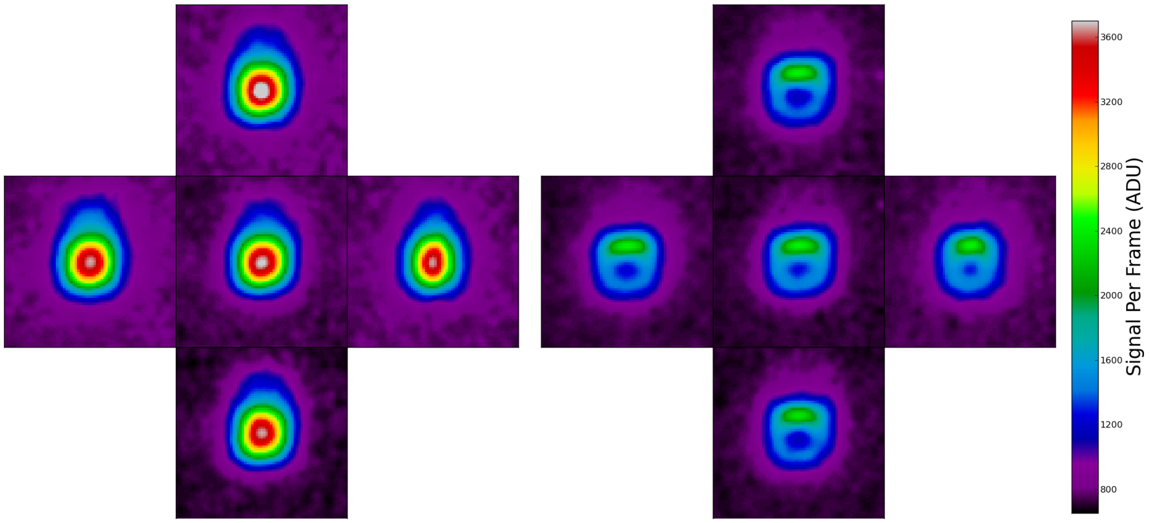


Figure C.21: PRF of a bad pixel along with the PRFs of adjacent pixels. The “bad” pixel from Figures C.7, C.19 and C.20 is shown. Left panels show pixels at low charge integration level (first SUR frame) and right panels show pixels at high charge integration level (sixth SUR frame).

My first and simplest theory involved charge leakage somewhat similar to capacitive coupling. This would entail a highly illuminated pixel “spilling” over its charge into adjacent pixels. To test this we plotted the low integrated signal and high integrated signal PRFs of pixels surrounding the “bad” pixel from Figure C.19. The results are shown in Figure C.21 with the first SUR frame on the left and the last SUR frame on the right. It is clear that the charge lost from the central pixel is not transferred to the adjacent ones and such charge spill-over cannot account for the observed behavior.

C.6.2 Readout Multiplexer

The readout channel structure could mean that the amplifiers mounted on the readout video cards are responsible. However the fact that pixel deformations increase and sensitivity decreases in regions with lower bias voltage that are not correlated with the readout scheme in any way seems to indicated that the readout multiplexer may be to blame. Furthermore, since the structure does not appear under uniform (or nearly uniform) illumination it is likely to also depend on signal *differences* between neighboring pixels.

Figure C.22 shows a simplified schematic of the detector multiplexer readout.

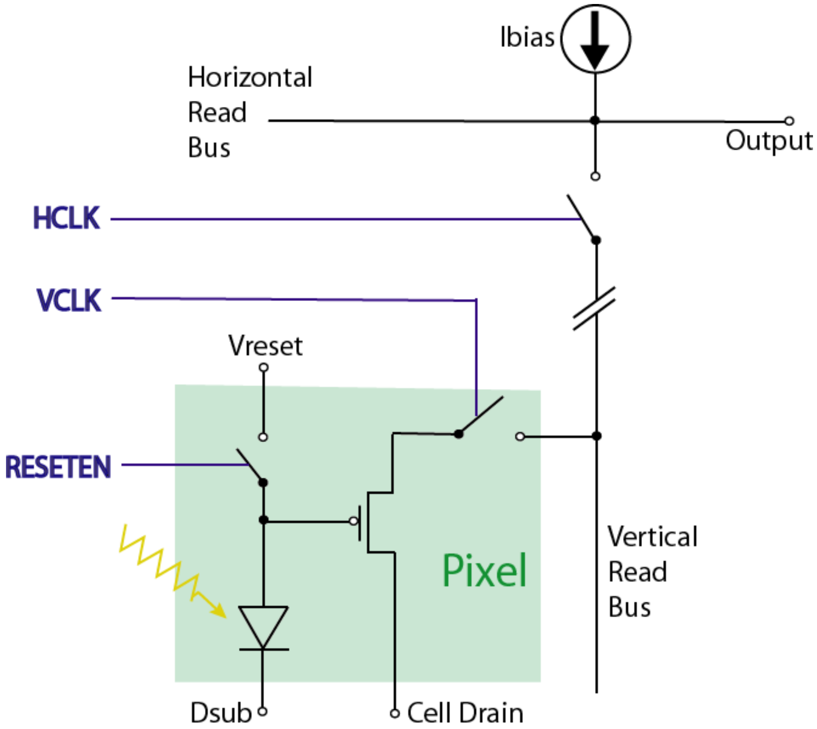


Figure C.22: The schematic of H2RG-236 readout.

When light falls onto the detector material (the diode in the schematic) it generates electron–hole pairs. The holes are collected at the pixel FET gate. For a FET that is turned on, the gate–source voltage V_{GS} is constant when the relatively constant current I_{bias} is turned on. Hence raising the gate voltage via the collected holes raises the source voltage by the same amount. This source voltage is then output as our signal. Since the reset voltage V_{reset} is kept constant, pixels with lower initial detector bias in Figure C.12 have a lower *natural* V_{GS} and therefore a lower drain–source voltage V_{DS}. These electrical differences in the FETs must therefore play a role in the effects we see. However they are not apparent on the right side of each readout channel. Hence the readout itself must be the primary cause. It should be noted that the read occurs from the right side towards the left.

The read sequence is as follows: all pixels within a row in a readout channel are connected to their respective *vertical read buses* due to the vertical clock $VCLK$. Then the horizontal clock $HCLK$ connects a single vertical read bus to the *horizontal read bus*. The bias current I_{bias} then flows through the pixel FET raising its source voltage by an amount equal to the gate voltage change. This source voltage is then amplified and read out. Subsequent $HCLK$s advance the read to the next vertical read bus and therefore, the next pixel. I believe that the vertical read bus may hold the key to the observed distortions since it could potentially serve as a means of coupling pixels within the same columns, specifically the currently read pixel to the pixel in the previously read row. Since the distortion exists primarily in the vertical direction, horizontal coupling is unlikely. The left–right magnitude of the distortion would then be related to when a pixel is read out relative to when its vertical read bus was connected; rightmost pixels are read out immediately after being connected to the bus while those further left have been connected for some period of time prior to being read out. This does not explain why the two leftmost pixels in a channel are relatively undistorted though. The exact mechanism remains unknown.

C.7 Discussion of Errors

Multiple systematic effects can impact our results. For instance, stage positioning may not be accurate or reproducible. To check potential errors CDS data was acquired in different modes; one where the spots would be moved horizontally one step, scanned vertically and then repeat (*vertical scan*), one where the spots would be moved vertically one step, scanned horizontally and then repeat (*horizontal scan*) and finally a *vertical scan* done where the scan directions are reversed (*reverse vertical scan*). Data from all of these scans was averaged for each pixel. In addition, each mode was averaged separately to see if they differed among each other. The left panel of Figure C.23 shows the difference between *vertical scan* and *horizontal scan* data for a sample pixel and the right panel shows the difference between *vertical scan* and *reverse vertical scan.* The averages where set to a maximum response of unity which means that a deviation of 0.01 indicates a difference equal to 1% of the maximum pixel response. Large deviations are not observed.

In addition, the stage motion is susceptible to small random offsets due to mechanical stress (bumping the optical table, coolant refill, etc). The data sets were examined and regions around obvious glitches were removed from the analysis. Scanning multiple spots over each pixel helps to average away remaining small stage motion errors. Unfortunately, our SUR data set included much fewer spots than our CDS set which lead to some pixels not being scanned at all.

Finally, the stage may have repeatable offsets built in due to its stepper motor design. Again, large statistical samples along with randomized scan start positions should allow us to average away this possible error. Figure C.24 shows the correlations between the pixel widths (left) and heights (right) computed using averages of *vertical scans*, *horizontal scans*, *reverse vertical scans* individually and an average of all scans.

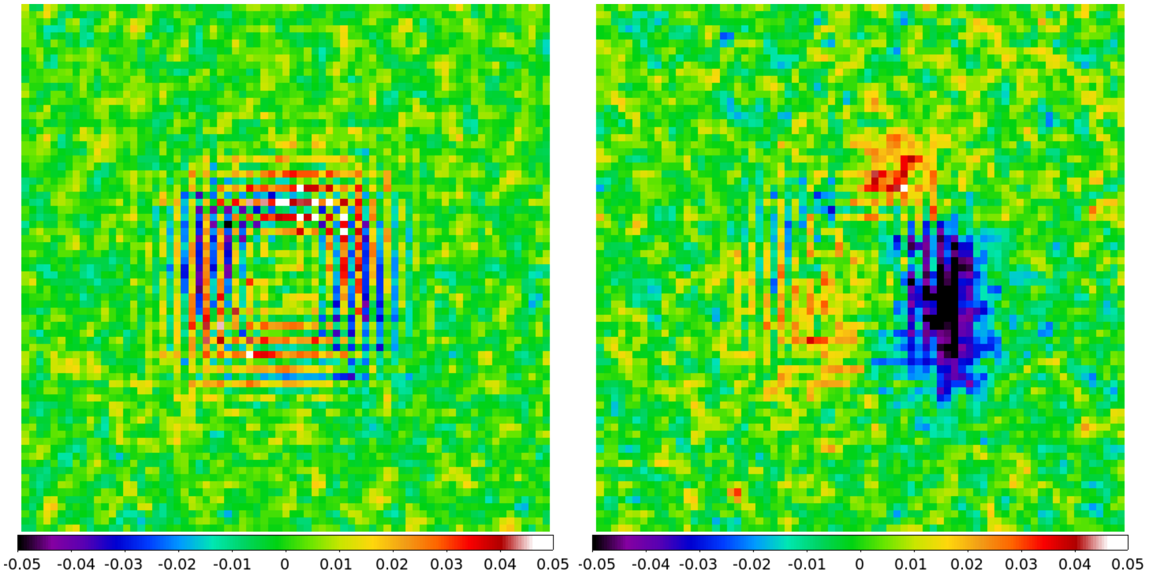


Figure C.23: Residuals between PRFs obtained with various stage scanning directions. Left panel: the difference between the *vertical scan* average PRF and the *horizontal scan* average PRF. Right panel: the difference between the *vertical scan* average PRF and the *reverse vertical scan* average PRF.

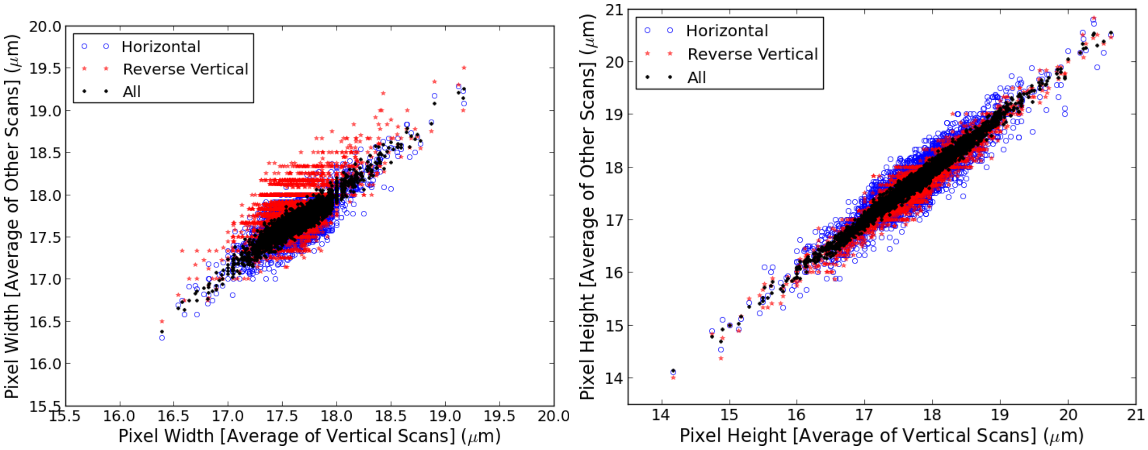


Figure C.24: Correlations between pixel sizes determined by averaging *vertical scans* and sizes determined by averaging just the *horizontal scans*, just the *reverse vertical scans* and all scans. Measured widths are on the left and heights are on the right.

As expected, *horizontal scans* constrain widths better than *vertical scans* and *vertical scans* constrain heights better. But, averaging multiple measurements shows that both the widths and heights recovered with either scan strategy agree and therefore the choice of strategy does not introduce systematics that cannot be overcome by a sufficiently large and randomized sample.

C.8 Summary

The Spots-O-Matic was designed primarily to improve photometry of point as well as extended sources. It has however revealed the presence of far more complicated structures that appear to be related to the electronics design, PSF size and total integrated signal. In addition, some large scale effective area variations have been detected that may be caused by mechanical stress experienced by the detector.

BIBLIOGRAPHY

Abbott, T., Abdalla, F., Achitouv, I., et al. 2012, The Astronomer's Telegram, 4668, 1

Abell, G. O. 1958, ApJS, 3, 211

Albert, L., Baril, M., Ward, J., Arnouts, S., & Devost, D. 2008, Proceedings of SPIE, 7014, 70142K

Aldering, G., Akerlof, C. W., Amanullah, R., et al. 2002, Proceedings of SPIE, 4835, 146

Amiaux, J., Scaramella, R., Mellier, Y., et al. 2012, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 8442, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Andersson, K., Benson, B. A., Ade, P. A. R., et al. 2011, The Astrophysical Journal, 738, 48

Angulo, R. E., Springel, V., White, S. D. M., et al. 2012, MNRAS, 426, 2046

Arnaud, M., Pratt, G. W., Piffaretti, R., et al. 2010, A&A, 517, A92

Baggett, S. M., Hill, R. J., Kimble, R. A., et al. 2008, in Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series, Vol. 7021, Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series

Baggett, S. M., Hill, R. J., Kimble, R. A., et al. 2008, Proceedings of SPIE, 7021, 70211Q

Bai, L., Marcillac, D., Rieke, G. H., et al. 2007, The Astrophysical Journal, 664, 181

Barron, N., Borysow, M., Beyerlein, K., et al. 2007, Publications of the Astronomical Society of the Pacific, 119, 466

Battaglia, N., Bond, J. R., Pfrommer, C., Sievers, J. L., & Sijacki, D. 2010, The Astrophysical Journal, 725, 91

Battye, R. A., & Weller, J. 2003, Phys. Rev. D, 68, 083506

Behroozi, P. S., Wechsler, R. H., & Wu, H.-Y. 2013, ApJ, 762, 109

Bennett, C. L., Halpern, M., Hinshaw, G., et al. 2003, ApJS, 148, 1

Benson, B. A., de Haan, T., Dudley, J. P., et al. 2013, The Astrophysical Journal, 763, 147

Biesiadzinski, T., Lorenzon, W., Newman, R., et al. 2011a, Publications of the Astronomical Society of the Pacific, 123, 958, ArticleType: research-article / Full publication date: August 2011 / Copyright 2011 The University of Chicago Press

—. 2011b, Publications of the Astronomical Society of the Pacific, 123, 179

Biesiadzinski, T., Lorenzon, W., Schubnell, M., Tarlé, G., & Weaverdyck, C. 2013, in Preparation. To be submitted to PASP

Biesiadzinski, T., McMahon, J., Miller, C. J., Nord, B., & Shaw, L. 2012, ApJ, 757, 1

Biesiadzinski, T. P., Tarlé, G., Howe, M. J., et al. 2010, in Proceedings of SPIE, Vol. 7742, 44

Birkinshaw, M. 1999, Phys. Rep., 310, 97

Biviano, A. 2000, in Constructing the Universe with Clusters of Galaxies

Bleem, L. E., van Engelen, A., Holder, G. P., et al. 2012, ApJ, 753, L9

Bode, P., & Ostriker, J. P. 2003, ApJS, 145, 1

Bode, P., Ostriker, J. P., & Vikhlinin, A. 2009, ApJ, 700, 989

Bohlin, R., Linder, D., & Riess, A. 2005, in NICMOS ISR, Vol. 2005-002 (Space Telescope Science Institute)

Böhringer, H., Voges, W., Huchra, J. P., et al. 2000, ApJS, 129, 435

Bonamente, M., Joy, M., LaRoque, S. J., et al. 2008, The Astrophysical Journal, 675, 106

Bonamente, M., Hasler, N., Bulbul, E., et al. 2012, New Journal of Physics, 14, 025010

Borgani, S., & Kravtsov, A. 2009, ArXiv e-prints, arXiv:0906.4370

Branch, D. 1998, ARA&A, 36, 17

Brown, M., Schubnell, M., & Tarlé, G. 2006, Publications of the Astronomical Society of the Pacific, 118, 1443, ArticleType: research-article / Full publication date: October 2006 / Copyright 2006 The University of Chicago Press

Brown, M. G. 2007, PhD thesis, University of Michigan

Budde, W. 1979, Applied Optics, 18, 1555

Burenin, R. A., Vikhlinin, A., Hornstrup, A., et al. 2007, ApJS, 172, 561

Busca, N. G., Delubac, T., Rich, J., et al. 2013, A&A, 552, A96

Busha, M., & et al. 2013, in Preparation

Carlstrom, J. E., Holder, G. P., & Reese, E. D. 2002, ARA&A, 40, 643

Carlstrom, J. E., Ade, P. A. R., Aird, K. A., et al. 2011, PASP, 123, 568

Cheng, L. 2009, Master's thesis, Chester F. Carlson Center for Imaging Science,Rochester Institute of Technology

Coble, K., Bonamente, M., Carlstrom, J. E., et al. 2007, The Astronomical Journal, 134, 897

Colless, M., Dalton, G., Maddox, S., et al. 2001, MNRAS, 328, 1039

Crommelin, A. C. D. 1918, JRASC, 12, 33

Cunha, C. 2009, Phys. Rev. D, 79, 063009

Das, S., Marriage, T. A., Ade, P. A. R., et al. 2011, ApJ, 729, 62

de Jong, R. S., Arribas, S., Barker, E. A., et al. 2006, in The 2005 HST Calibration Workshop: Hubble After the Transition to Two-Gyro Mode, ed. A. M. Koekemoer, P. Goudfrooij, & L. L. Dressel, 121

De Zotti, G., Ricci, R., Mesa, D., et al. 2005, A&A, 431, 893

Delabrouille, J., Betoule, M., Melin, J.-B., et al. 2013, A&A, 553, A96

Deustua, S., Long, K. S., McCullough, P., et al. 2010, Proceedings of SPIE, 7731, 77313C

Dicke, R. H., Peebles, P. J. E., Roll, P. G., & Wilkinson, D. T. 1965, ApJ, 142, 414

Draper, P., Dodelson, S., Hao, J., & Rozo, E. 2012, Phys. Rev. D, 85, 023005

Dressler, A., & Shectman, S. A. 1987, AJ, 94, 899

Duffy, A. R., Schaye, J., Kay, S. T., & Dalla Vecchia, C. 2008, MNRAS, 390, L64

Dunkley, J., Komatsu, E., Nolta, M. R., et al. 2009, ApJS, 180, 306

Dvali, G., Gabadadze, G., & Porrati, M. 2000, Physics Letters B, 485, 208

Efstathiou, G., Sutherland, W. J., & Maddox, S. J. 1990, Nature, 348, 705

Einstein, A. 1915, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847., 844

—. 1917, Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152., 142

Eisenstein, D. J., Zehavi, I., Hogg, D. W., et al. 2005, ApJ, 633, 560

ESA, & Carreau, C. 2013, The history of structure formation in the Universe

Evrard, A. E., MacFarland, T. J., Couchman, H. M. P., et al. 2002, ApJ, 573, 7

Finger, G., Beletic, J. W., Dorn, R., et al. 2005, Experimental Astronomy, 19, 135

Finger, G., Dorn, R. J., Meyer, M., et al. 2004, in Proceedings of SPIE, Vol. 5499, 47–58

Firester, A. H., Heller, M. E., & Sheng, P. 1977, Applied Optics, 16, 1971

Fixsen, D. J., Cheng, E. S., Gales, J. M., et al. 1996, ApJ, 473, 576

Flaugher, B. 2011, dES Collaboration Meeting, Portsmouth

Friedman, A. 1922, Zeitschrift fr Physik, 10, 377

Friedmann, A. 1924, Zeitschrift fr Physik, 21, 326

Frieman, J. A., Turner, M. S., & Huterer, D. 2008, ARA&A, 46, 385

Gal, R. R. 2006, ArXiv Astrophysics e-prints, arXiv:astro-ph/0601195

Geller, M. J., & Huchra, J. P. 1989, Science, 246, 897

Goldhaber, G. 2009, in American Institute of Physics Conference Series, Vol. 1166, American Institute of Physics Conference Series, ed. D. B. Cline, 53–72

Górski, K. M., Hivon, E., Banday, A. J., et al. 2005, ApJ, 622, 759

Gott, III, J. R., Jurić, M., Schlegel, D., et al. 2005, ApJ, 624, 463

Granato, G. L., Zotti, G. D., Silva, L., Bressan, A., & Danese, L. 2004, The Astrophysical Journal, 600, 580

Gruppioni, C., Lari, C., Pozzi, F., et al. 2002, Monthly Notices of the Royal Astronomical Society, 335, 831

Günster, S., Ristau, D., & Davies, R. 2011, in Proceedings of SPIE, Vol. 8168, 81681Y–81681Y–12

Guth, A. H. 1981, Phys. Rev. D, 23, 347

Guth, A. H., & Pi, S.-Y. 1982, Physical Review Letters, 49, 1110

Haiman, Z., Mohr, J. J., & Holder, G. P. 2001, ApJ, 553, 545

Hashimoto, Y., Augustus Oemler, J., Lin, H., & Tucker, D. L. 1998, The Astrophysical Journal, 499, 589

Herranz, D., Sanz, J. L., Hobson, M. P., et al. 2002, MNRAS, 336, 1057

Hill, R., Malumuth, E., Foltz, R., et al. 2009, in Detectors for Astronomy (ESO Garching Workshop on Detectors for Astronomy)

Hill, R. J., Malumuth, E., Foltz, R., et al. 2010, in Proceedings of SPIE, Vol. 7742, 56

Hinshaw, G., Larson, D., Komatsu, E., et al. 2012, ArXiv e-prints, arXiv:1212.5226

Hogg, D. W., Eisenstein, D. J., Blanton, M. R., et al. 2005, ApJ, 624, 54

Holder, G., Haiman, Z., & Mohr, J. J. 2001, ApJ, 560, L111

Hubble, E. 1925a, Contributions from the Mount Wilson Observatory / Carnegie Institution of Washington, 304, 1

—. 1929, Proceedings of the National Academy of Science, 15, 168

Hubble, E. P. 1925b, The Observatory, 48, 139

Huterer, D., & Linder, E. V. 2007, Phys. Rev. D, 75, 023519

Ivezic, Z., Tyson, J. A., Acosta, E., et al. 2008, ArXiv e-prints, arXiv:0805.2366

Janesick, J., Klaasen, K., & Elliott, T. 1985, Proceedings of SPIE, 570, 7

Johnston, D. E., Sheldon, E. S., Wechsler, R. H., et al. 2007, ArXiv e-prints, arXiv:0709.1159

Kaiser, N. 1986, MNRAS, 222, 323

—. 1991, ApJ, 383, 104

Kay, S. T., Peel, M. W., Short, C. J., et al. 2012, Monthly Notices of the Royal Astronomical Society, 422, 1999

Keisler, R., Reichardt, C. L., Aird, K. A., et al. 2011, ApJ, 743, 28

Kessler, R., Becker, A. C., Cinabro, D., et al. 2009, ApJS, 185, 32

Kim, A. G., Linder, E. V., Miquel, R., & Mostek, N. 2004, MNRAS, 347, 909

Koester, B. P., McKay, T. A., Annis, J., et al. 2007a, ApJ, 660, 239

—. 2007b, ApJ, 660, 221

Kolb, E. W., & Turner, M. S. 1990, The early universe. (Westview Press)

Komatsu, E., Matsuo, H., Kitayama, T., et al. 2001, PASJ, 53, 57

Kravtsov, A. V., & Borgani, S. 2012, ARA&A, 50, 353

Kravtsov, A. V., Vikhlinin, A., & Nagai, D. 2006, ApJ, 650, 128

Leach, R. W., & Low, F. J. 2000a, Proceedings of SPIE, 4008, 337

Leach, R. W., & Low, F. J. 2000b, in Proceedings of SPIE, Vol. 4008, 337–343

Lemaître, G. 1931, MNRAS, 91, 483

Lewis, A., Challinor, A., & Lasenby, A. 2000, ApJ, 538, 473

Lima, M., & Hu, W. 2004, Phys. Rev. D, 70, 043504

Lima, M., Jain, B., & Devlin, M. 2010a, Monthly Notices of the Royal Astronomical Society, 406, 2352

Lima, M., Jain, B., Devlin, M., & Aguirre, J. 2010b, The Astrophysical Journal Letters, 717, L31

Linder, E. V. 2005, Phys. Rev. D, 72, 043529

Lorenzon, W., Newman, R., Schubnell, M., Tarlé, G., & Weaverdyck, C. 2008, in Proceedings of SPIE, Vol. 7021, 70210V–70210V–7

Maddox, S. J., Efstathiou, G., Sutherland, W. J., & Loveday, J. 1990, MNRAS, 242, 43P

Mahdavi, A., Hoekstra, H., Babul, A., et al. 2013, ApJ, 767, 116

Marriage, T. A., Acquaviva, V., Ade, P. A. R., et al. 2011, ApJ, 737, 61

Marrone, D. P., Smith, G. P., Okabe, N., et al. 2012, The Astrophysical Journal, 754, 119

McCullough, P. R., Regan, M., Bergeron, L., & Lindsay, K. 2008, Publications of the Astronomical Society of the Pacific, 120, 759

McMahon, R. 2010, in DES-DocDB

McMahon, R. 2012, in Science from the Next Generation Imaging and Spectroscopic Surveys

McNamara, B. R., & Nulsen, P. E. J. 2007, ARA&A, 45, 117

Melin, J.-B., Bartlett, J. G., & Delabrouille, J. 2006, A&A, 459, 341

Miller, C. J., Nichol, R. C., Reichart, D., et al. 2005, AJ, 130, 968

Mohr, J. J. 2005, in Astronomical Society of the Pacific Conference Series, Vol. 339, Observing Dark Energy, ed. S. C. Wolff & T. R. Lauer, 140

Molnar, S. M., Haiman, Z., Birkinshaw, M., & Mushotzky, R. F. 2004, ApJ, 601, 22

Moore, A. C. 2006, Optical Engineering, 45, 076402

Moore, A. C., Ninkov, Z., & Forrest, W. J. 2004, Proceedings of SPIE, 5167, 204

Mortara, L., & Fowler, A. 1981, Proceedings of SPIE, 290, 28

Mortonson, M. J., Hu, W., & Huterer, D. 2011, Phys. Rev. D, 83, 023015

Mostek, N. 2007, PhD thesis, Indiana University

Motl, P. M., Hallman, E. J., Burns, J. O., & Norman, M. L. 2005, ApJ, 623, L63

Nagai, D. 2006, ApJ, 650, 538

Nagai, D., Kravtsov, A. V., & Vikhlinin, A. 2007, The Astrophysical Journal, 668, 1

Navarro, J. F., Frenk, C. S., & White, S. D. M. 1996, ApJ, 462, 563

Negrello, M., Perrotta, F., Gonzlez, J. G.-N., et al. 2007, Monthly Notices of the Royal Astronomical Society, 377, 1557

Nord, B. D. 2010, PhD thesis, University of Michigan

Norton, P. 2002, Opto-electronics review, 10, 159

O'Dea, D. T., Clark, C. N., Contaldi, C. R., & MacTavish, C. J. 2012, Monthly Notices of the Royal Astronomical Society, 419, 1795

Penzias, A. A., & Wilson, R. W. 1965, ApJ, 142, 419

Percival, W. J., Reid, B. A., Eisenstein, D. J., et al. 2010, MNRAS, 401, 2148

Perlmutter, S., Aldering, G., Deustua, S., et al. 1997, in Bulletin of the American Astronomical Society, Vol. 29, American Astronomical Society Meeting Abstracts, 1351

Perlmutter, S., Aldering, G., Goldhaber, G., et al. 1999, ApJ, 517, 565

Piffaretti, R., Arnaud, M., Pratt, G. W., Pointecouteau, E., & Melin, J.-B. 2011, A&A, 534, A109

Piffaretti, R., & Valdarnini, R. 2008, A&A, 491, 71

Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2011a, A&A, 536, A9

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2011b, A&A, 536, A8

—. 2011c, A&A, 536, A11

Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2011d, A&A, 536, A12

Planck Collaboration, Aghanim, N., Arnaud, M., et al. 2012, A&A, 543, A102

Planck Collaboration, Ade, P. A. R., Aghanim, N., et al. 2013a, ArXiv e-prints, arXiv:1303.5062

—. 2013b, ArXiv e-prints, arXiv:1303.5076

—. 2013c, ArXiv e-prints, arXiv:1303.5083

—. 2013d, A&A, 550, A129

—. 2013e, A&A, 550, A130

Planck HFI Core Team, Ade, P. A. R., Aghanim, N., et al. 2011, A&A, 536, A6

Press, W. H., & Schechter, P. 1974, ApJ, 187, 425

Rasia, E., Mazzotta, P., Evrard, A., et al. 2011, ApJ, 729, 45

Rasia, E., Meneghetti, M., Martino, R., et al. 2012, New Journal of Physics, 14, 055018

Reichardt, C. L., Stalder, B., Bleem, L. E., et al. 2013, ApJ, 763, 127

Reid, B. A., Percival, W. J., Eisenstein, D. J., et al. 2010, MNRAS, 404, 60

Riess, A. 2010, in WFC3 ISR, Vol. 2010-07 (Space Telescope Science Institute)

Riess, A. G., Filippenko, A. V., Challis, P., et al. 1998, AJ, 116, 1009

Riess, A. G., Macri, L., Casertano, S., et al. 2009, ApJ, 699, 539

Robertson, H. P. 1935, ApJ, 82, 284

Rozo, E., Bartlett, J. G., Evrard, A. E., & Rykoff, E. S. 2012, ArXiv e-prints, arXiv:1204.6305

Rozo, E., Rykoff, E. S., Evrard, A., et al. 2009a, ApJ, 699, 768

Rozo, E., Rykoff, E. S., Koester, B. P., et al. 2009b, The Astrophysical Journal, 703, 601

Rozo, E., Wechsler, R. H., Rykoff, E. S., et al. 2010, ApJ, 708, 645

Rubin, V. C., & Ford, Jr., W. K. 1970, ApJ, 159, 379

Rykoff, E. S., McKay, T. A., Becker, M. R., et al. 2008, ApJ, 675, 1106

Sanderson, A. J. R., Edge, A. C., & Smith, G. P. 2009, MNRAS, 398, 1698

Sazonov, S. Y., & Sunyaev, R. A. 1999, MNRAS, 310, 765

Schaffer, K. K., Crawford, T. M., Aird, K. A., et al. 2011, The Astrophysical Journal, 743, 90

Schubnell, M., Biesiadzinski, T., Landes, B., et al. 2009, Proceedings of Detectors for Astronomy

Schubnell, M., Biesiadzinski, T., Lorenzon, W., Newman, R., & Tarlé, G. 2010, in Proceedings of SPIE, Vol. 7742, 45

Schubnell, M., Brown, M. G., Karabina, A., et al. 2008, in Proceedings of SPIE, Vol. 7021, 17

Schubnell, M., Barron, N., Bebek, C., et al. 2006, Proceedings of SPIE, 6276, 62760Q

Sehgal, N., Bode, P., Das, S., et al. 2010, The Astrophysical Journal, 709, 920

Sehgal, N., Trac, H., Acquaviva, V., et al. 2011, ApJ, 732, 44

Shaw, L. D., Holder, G. P., & Bode, P. 2008, ApJ, 686, 206

Sheldon, E. S., Johnston, D. E., Scranton, R., et al. 2009, ApJ, 703, 2217

Sievers, J. L., Hlozek, R. A., Nolta, M. R., et al. 2013, ArXiv e-prints, arXiv:1301.0824

Sifón, C., Menanteau, F., Hasselfield, M., et al. 2013, ApJ, 772, 25

Silva, L., De Zotti, G., Granato, G. L., Maiolino, R., & Danese, L. 2004, ArXiv Astrophysics e-prints, arXiv:astro-ph/0403166

Smith, R. 2010, internal Presentation

Smith, R. M., & Rahmer, G. 2008, in Proceedings of SPIE, Vol. 7021, 70212A–70212A–12

Smith, R. M., Zavodny, M., Rahmer, G., & Bonati, M. 2008, Proceedings of SPIE, 7021, 70210J

Smoot, G. F., Bennett, C. L., Kogut, A., et al. 1992, ApJ, 396, L1

SNAP Collaboration. 2005, ArXiv Astrophysics e-prints, arXiv:astro-ph/0507460

Spergel, D., Gehrels, N., Breckinridge, J., et al. 2013, ArXiv e-prints, arXiv:1305.5425

Spergel, D. N., Verde, L., Peiris, H. V., et al. 2003, ApJS, 148, 175

Springel, V., & et al. 2005, Nature, 435, 629636

Springel, V., White, M., & Hernquist, L. 2001, ApJ, 549, 681

Springel, V., White, S. D. M., & Frenk, C. S. 2006, Nature, 440, 1137

Stanek, R., Rasia, E., Evrard, A. E., Pearce, F., & Gazzola, L. 2010, The Astrophysical Journal, 715, 1508

Stanek, R. M. 2009, PhD thesis, University of Michigan

Staniszewski, Z., Ade, P. A. R., Aird, K. A., et al. 2009, ApJ, 701, 32

Steigman, G. 2007, Annual Review of Nuclear and Particle Science, 57, 463

Sunyaev, R. A., & Zeldovich, Y. B. 1972, Comments on Astrophysics and Space Physics, 4, 173

Swetz, D. S., Ade, P. A. R., Amiri, M., et al. 2011, ApJS, 194, 41

The Dark Energy Survey Collaboration. 2005, ArXiv Astrophysics e-prints, arXiv:astro-ph/0510346

Vanderlinde, K., Crawford, T. M., de Haan, T., et al. 2010, ApJ, 722, 1180

Vieira, J. D., Crawford, T. M., Switzer, E. R., et al. 2010, The Astrophysical Journal, 719, 763

Vikhlinin, A., Burenin, R. A., Ebeling, H., et al. 2009, The Astrophysical Journal, 692, 1033

Vikhlinin, A., Kravtsov, A. V., Burenin, R. A., et al. 2009, ApJ, 692, 1060

Voges, W., Aschenbach, B., Boller, T., et al. 1999, A&A, 349, 389

Voit, G. M. 2005, Reviews of Modern Physics, 77, 207

Walker, A. G. 1937, Proceedings of the London Mathematical Society, s2-42, 90

Wang, L., & Steinhardt, P. J. 1998, ApJ, 508, 483

Wang, S., Khoury, J., Haiman, Z., & May, M. 2004, Phys. Rev. D, 70, 123008

Wechsler, R., & et al. 2013, in Preparation

Weinberg, D., Bard, D., Dawson, K., et al. 2013, in Snowmass CF5 Papers (Snowmass on the Mississippi (CSS 2013))

Weinberg, D. H., Mortonson, M. J., Eisenstein, D. J., et al. 2012, ArXiv e-prints, arXiv:1201.2434

Weinberg, S. 2008, Cosmology (New York, New York: Oxford University Press)

White, M., Cohn, J. D., & Smit, R. 2010, MNRAS, 408, 1818

White, M., Hernquist, L., & Springel, V. 2002, ApJ, 579, 16

White, S. D. M., Navarro, J. F., Evrard, A. E., & Frenk, C. S. 1993, Nature, 366, 429

Williamson, R., Benson, B. A., High, F. W., et al. 2011, ApJ, 738, 139

Wu, H.-Y., Rozo, E., & Wechsler, R. H. 2010, ApJ, 713, 1207

Yoon, H. W., Butler, J. J., Larason, T. C., & Eppeldauer, G. P. 2003, Metrologia, 40, S154

Zentner, A. R. 2007, International Journal of Modern Physics D, 16, 763

Zwicky, F. 1937, ApJ, 86, 217